

Data-driven model development in environmental geography

Methodological advancements and scientific
applications

kumulative
Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften
(Dr. rer. nat.)

dem Fachbereich Geographie
der Philipps-Universität Marburg
vorgelegt von

Hanna Meyer
aus Oberhausen

Marburg, Juli 2017

Am Fachbereich Geographie der Philipps-Universität Marburg
(Hochschulkennziffer: 1180) am 17.07.2017 als Dissertation eingereicht.

Erstgutachter: Prof. Dr. Thomas Nauß
Zweitgutachter: Prof. Dr. Bernhard Seeger
Drittgutachter: Prof. Dr. Andreas Huth

Tag der mündlichen Prüfung: 08.02.2018

Preface

At the end of a challenging but yet inspiring and special time, I would like to thank those people who supported and accompanied me on this journey. First of all, I would like to sincerely thank Thomas Nauß for providing the opportunity to write this thesis, his great support throughout the entire process and eventually for always allowing more than just a few (spatial and scientific) detours. Further, I would like to express my deep gratitude to Chris Reudenbach for his support in so many ways, that I could always count on and that I appreciate a lot.

I am also very grateful to my colleagues in the working group of environmental informatics. The time would not have been as special without them as they provided a great working (and beyond) environment that is certainly unique. As the first steps are always the hardest, I am especially thankful to Meike Kühnlein for introducing me to the field of rainfall modelling. Further, thanks go to Alice Ziegler for proofreading the envelope of this thesis and her valuable comments.

However, I would not have been there without the early support of Jörg Bendix and Boris Thies who were responsible that I could acquire a taste in science during my studies. In this context, I would further like to thank Lukas Lehnert for many years of fruitful discussions, sharing knowledge, ideas and field experiences.

This thesis was embedded in the BMBF funded IDESSA project and would not have been possible without this financial support. It further relied on data from several institutions that I would like to acknowledge at the end of each respective chapter of my thesis. The papers of this thesis were developed in collaboration with several people. I would like to thank all co-authors for their contributions and I am grateful to Caley Gasch, Tom Hengl, Karoline Messenzehl and Marwan Katurji for giving me the chance to contribute to their projects. Special thanks go to Marwan Katurji and Peyman Zawar-Reza for welcoming me in New Zealand several times and for introducing me to the Antarctic research. I am also thankful to the IDESSA team for scientific discussions and help during field work.

Last but not least, the thesis could not have been developed without the general support of my family and friends. I am very thankful to my neighbours Stefan, Anneke and Lotta for distraction after work and to Nic, for patiently proofreading my papers and much more importantly for being an awesome partner for mountain adventures in New Zealand and beyond. Finally, very special thanks go to my sister Nele for never-ending scientific discussions and especially for being an invaluable mate in life.

Hanna Meyer
Marburg, July 2017

Abstract

One key task in environmental geography is obtaining information of geographic features in space or in space and time. For this purpose, modelling strategies are needed that allow a delineation of spatio-temporal information based on limited field data. In this context, the nonlinearity and complexity of environmental systems require modelling strategies that allow handling arbitrary relationships and large sets of potential predictor variables. These requirements provoke a paradigm-shift from a parametric towards a non-parametric and data-driven model development which is strengthened by an increasing availability of geographic data. In that respect, machine learning algorithms have been proven to be an important tool to learn patterns in nonlinear and complex systems. While the large number of machine learning applications in scientific journals as well as recent software developments nowadays feign a simplicity of these methods, their application is not a trivial task. This holds especially true for geographic data as they have certain characteristics, especially spatial dependency, that make them stand out against the mass of "ordinary" data. However, this is widely ignored in geographic machine learning applications.

This thesis assesses the potential and the sensitivity of machine learning in environmental geography. In this context, a number of machine learning applications in a broad spectrum of environmental geography have been published, providing a collection of comprehensive knowledge about machine learning in environmental geography. The individual contributions are incorporated in the major hypothesis that, only if characteristics of geospatial data are considered, data-driven modelling strategies lead to a reliable gain of information and to robust spatio-temporal model results. Beside this superior methodological focus, each application aims at providing new insights in its respective field of research.

In this thesis, a number of relevant environmental monitoring products have been developed. The results emphasize that a high expertise of the machine learning methods as well as of the scientific field is crucial to advance the environmental geography. The thesis is the first to raise awareness of spatial or spatio-temporal over-fitting in geographic machine learning applications and the significant consequences to the outcome. To approach this problem, a new method for model development is provided that is adapted for geographic data and allows for improved model results. The thesis is finally an appeal to think beyond the "standard machine learning way" as it proves that applying standard machine learning concepts on geographic data results in considerable over-fitting and misinterpretation of the results. Only when characteristics of geographic data are considered, machine learning provides a powerful tool to provide scientifically valuable results in environmental geography.

Zusammenfassung

Die Erfassung räumlich kontinuierlicher Daten und raum-zeitlicher Dynamiken ist ein Forschungsschwerpunkt der Umweltgeographie. Zu diesem Ziel sind Modellierungsmethoden erforderlich, die es ermöglichen, aus limitierten Felddaten raum-zeitliche Aussagen abzuleiten. Die Komplexität von Umweltsystemen erfordert dabei die Verwendung von Modellierungsstrategien, die es erlauben, beliebige Zusammenhänge zwischen einer Vielzahl potentieller Prädiktoren zu berücksichtigen. Diese Anforderung verlangt nach einem Paradigmenwechsel von der parametrischen hin zu einer nicht-parametrischen, datengetriebenen Modellentwicklung, was zusätzlich durch die zunehmende Verfügbarkeit von Geodaten verstärkt wird. In diesem Zusammenhang haben sich maschinelle Lernverfahren als ein wichtiges Werkzeug erwiesen, um Muster in nicht-linearen und komplexen Systemen zu erfassen. Durch die wachsende Popularität maschineller Lernverfahren in wissenschaftlichen Zeitschriften und die Entwicklung komfortabler Softwarepakete wird zunehmend der Fehleindruck einer einfachen Anwendbarkeit erzeugt. Dem gegenüber steht jedoch eine Komplexität, die im Detail nur durch eine umfassende Methodenkompetenz kontrolliert werden kann. Diese Problematik gilt insbesondere für Geodaten, die besondere Merkmale wie vor allem räumliche Abhängigkeit aufweisen, womit sie sich von "gewöhnlichen" Daten abheben, was jedoch in maschinellen Lernanwendungen bisher weitestgehend ignoriert wird.

Die vorliegende Arbeit beschäftigt sich mit dem Potenzial und der Sensitivität des maschinellen Lernens in der Umweltgeographie. In diesem Zusammenhang wurde eine Reihe von maschinellen Lernanwendungen in einem breiten Spektrum der Umweltgeographie veröffentlicht. Die einzelnen Beiträge stehen unter der übergeordneten Hypothese, dass datengetriebene Modellierungsstrategien nur dann zu einem Informationsgewinn und zu robusten raum-zeitlichen Ergebnissen führen, wenn die Merkmale von geographischen Daten berücksichtigt werden. Neben diesem übergeordneten methodischen Fokus zielt jede Anwendung darauf ab, durch adäquat angewandte Methoden neue fachliche Erkenntnisse in ihrem jeweiligen Forschungsgebiet zu liefern.

Im Rahmen der Arbeit wurde eine Vielzahl relevanter Umweltmonitoring-Produkte entwickelt. Die Ergebnisse verdeutlichen, dass sowohl hohe fachwissenschaftliche als auch methodische Kenntnisse unverzichtbar sind, um den Bereich der datengetriebenen Umweltgeographie voranzutreiben. Die Arbeit demonstriert erstmals die Relevanz räumlicher Überfittung in geographischen Lernanwendungen und legt ihre Auswirkungen auf die Modellergebnisse dar. Um diesem Problem entgegenzuwirken, wird eine neue, an Geodaten angepasste Methode zur Modellentwicklung entwickelt, wodurch deutlich verbesserte Ergebnisse erzielt

werden können. Diese Arbeit ist abschließend als Appell zu verstehen, über die Standardanwendungen der maschinellen Lernverfahren hinauszudenken, da sie beweist, dass die Anwendung von Standardverfahren auf Geodaten zu starker Überfittung und Fehlinterpretation der Ergebnisse führt. Erst wenn Eigenschaften von geographischen Daten berücksichtigt werden, bietet das maschinelle Lernen ein leistungsstarkes Werkzeug, um wissenschaftlich verlässliche Ergebnisse für die Umweltgeographie zu liefern.

Contents

Preface	I
Abstract	III
Zusammenfassung	V
List of Figures	XIII
List of Tables	XVII
List of Acronyms	XIX
1 Introduction	2
1.1 Modeling tasks in environmental geography	2
1.2 Predictive modelling strategies	4
1.2.1 Classic modelling strategies	5
1.2.1.1 Spatial Interpolations	5
1.2.1.2 Statistical parametric models	6
1.2.2 The paradigm shift towards data-driven model development	7
1.3 Machine learning fundamentals and terminology	8
1.4 Machine learning in environmental geography - State of the art . .	9
1.4.1 Biogeography	10
1.4.2 Climatology	10
1.4.3 Soil science and hydrology	11
1.4.4 Geomorphology	11
1.5 Formulation of the scientific problem, aims and hypotheses	12
1.5.1 Characteristics of geographic data	12
1.5.2 Aims and hypotheses	13
1.5.3 Concept and structure of this thesis	14
2 Comparison of ML algorithms for rainfall retrievals	18
2.1 Introduction	19
2.2 Data and methodology	20
2.2.1 Datasets	21
2.2.1.1 Satellite data	21
2.2.1.2 RADOLAN RW data	22
2.2.2 Preprocessing of SEVIRI and Radar data	22
2.2.3 Predictor variables	23
2.2.4 Compilation of training and testing datasets	24

2.2.5	Model tuning	25
2.2.5.1	Random forests tuning	25
2.2.5.2	NNET and AVNNET tuning	26
2.2.5.3	SVM tuning	26
2.2.5.4	Threshold tuning for rainfall area classification models	26
2.2.6	Model prediction and validation	27
2.2.6.1	Validation of rainfall area classification models	27
2.2.6.2	Validation of rainfall rates regression models	28
2.3	Results	29
2.3.1	Comparison of predicted rainfall areas	29
2.3.2	Comparison of predicted rainfall rates	30
2.4	Discussion	30
2.4.1	Prediction of rainfall areas	30
2.4.2	Prediction of rainfall rates	32
2.4.3	Technical considerations	35
2.5	Summary and conclusions	36
3	Spectral and textural predictors for rainfall retrievals	40
3.1	Introduction	41
3.2	Methods	42
3.2.1	Satellite and ground truth data	43
3.2.2	Compilation of training and test data sets	44
3.2.3	Neural network training	45
3.2.3.1	Recursive feature selection	45
3.2.3.2	Fine tuning and model training	46
3.3	Results	46
3.4	Discussion and Conclusion	47
4	Satellite-based mapping of rainfall over southern Africa	52
4.1	Introduction	53
4.2	Methods	54
4.2.1	Study area	55
4.2.2	Data and Preprocessing	56
4.2.2.1	Station data	56
4.2.2.2	Satellite data	56
4.2.2.3	Cloud mask	57
4.2.3	Model strategies for rainfall estimation	57
4.2.3.1	General model framework	57
4.2.3.2	Training and test data sets	58

4.2.3.3	Tuning and model training	59
4.2.3.4	Spatial estimations of rainfall	60
4.2.4	Validation	60
4.2.5	Comparison to GPM	61
4.3	Results	62
4.3.1	Model performance	62
4.3.2	Comparison to GPM	63
4.4	Discussion	65
4.5	Conclusions	69
5	Mapping daily air temperature for Antarctica	72
5.1	Introduction	73
5.2	Methods	74
5.2.1	Data and Preprocessing	74
5.2.1.1	LST	74
5.2.1.2	Station Records	75
5.2.1.3	Auxiliary Data	75
5.2.1.4	Compilation of Model Training and Testing Data	76
5.2.2	Modelling	76
5.2.2.1	Algorithms	76
5.2.2.2	Cross-Validation Strategies and Feature Selection to Minimize Overfitting	77
5.2.2.3	Final Model Training, Evaluation and Prediction .	78
5.3	Results	79
5.3.1	Selected Features	79
5.3.2	Model Comparison and Evaluation	80
5.4	Discussion	86
5.5	Conclusions	88
6	From spectral measurements to maps of pasture degradation	92
6.1	Introduction	93
6.2	Data and methods	94
6.2.1	Study area	94
6.2.2	Field work	95
6.2.2.1	Spectral measurements	96
6.2.2.2	Proxies for pasture degradation	97
6.2.3	Satellite data	97
6.2.4	Methodology to remotely derive pasture degradation pa- rameters	97
6.2.4.1	Calculation of predictor feature spaces	98

6.2.4.2	Estimation of pasture degradation parameters using Random Forests	99
6.2.4.3	Regionalization	100
6.3	Results	101
6.3.1	Feature selection and predictive importance of NBIs	101
6.3.2	Accuracies of estimations based on hyper- and multispectral data	101
6.3.3	Application to satellite data	103
6.4	Discussion	105
6.5	Conclusions	109
7	Classifying woody vegetation in Google Earth images	112
7.1	Introduction	113
7.2	Methods	114
7.2.1	Study area	115
7.2.2	Data and variables	115
7.2.3	Random Forests classification	117
7.2.4	Validation	117
7.2.5	Prediction on new Google Earth images	118
7.3	Results and Discussion	119
7.3.1	Model performance and variable importance	119
7.3.2	Database of training sites	120
7.4	Conclusions	123
8	Spatio-temporal interpolation of soil properties in 3D+T	126
8.1	Introduction	127
8.2	Materials and methods	129
8.2.1	The Cook Agronomy Farm data set	129
8.2.2	Conceptual foundation for 3D+T modeling	131
8.2.3	3D+T random forests model	132
8.2.4	3D+T kriging model	134
8.2.5	Cross-validation	138
8.2.6	Software implementation	139
8.3	Results	140
8.3.1	3D+T random forests model	140
8.3.2	3D+T kriging model	140
8.3.3	Model accuracy	142
8.4	Discussion	144
8.4.1	Model performance	144
8.4.2	Interpretation of model predictions	149

8.4.3	Final conclusions and future directions	149
9	Soil respiration based on mid-infrared spectroscopy	154
9.1	Introduction	155
9.2	Material and Methods	156
9.2.1	Study area	156
9.2.2	Soil sampling	157
9.2.3	Determination of physicochemical soil properties	158
9.2.4	Soil respiration measurements and determination of Q10	158
9.2.5	Mid infrared spectroscopy (MIRS)	160
9.2.6	Relation between MIRS spectra and soil respiration parameters	160
9.2.7	Partial least square regression (PLSR)	160
9.2.8	Random Forest modeling	161
9.3	Results and Discussions	162
9.3.1	MIRS-PLSR based prediction of soil respiration at standardized temperature and moisture	162
9.3.2	MIRS-PLSR based prediction of Q10 values	163
9.3.3	Simultaneous prediction of soil respiration across various levels of soil moisture and temperature by Random Forest modeling	166
9.4	Conclusion	170
10	Regional-scale controls of rockfalls	172
10.1	Introduction	174
10.2	Characteristics of the study area	177
10.3	Modelling approach	178
10.3.1	Data selection and pre-processing	178
10.3.1.1	Response variable	178
10.3.1.2	Predictor variables	179
10.3.2	Validation methodology	182
10.3.3	Principal component analysis and logistic regression modelling	183
10.3.4	Random forest model	185
10.4	Results	186
10.4.1	Spatial characteristics of rockfall source areas (rockfall density statistics)	186
10.4.2	Principal components and geomorphic meaning	187
10.4.3	PCLR model and importance of PCs	188
10.4.4	Random forest model and variable importance	191
10.4.5	Validation and evaluation of model performances	191

10.5 Discussion	193
10.5.1 Evaluation of the methodological approach	193
10.5.2 Regional-scale controls on rockfall activity	195
10.5.2.1 The predisposing effect of rock mechanical characteristics	195
10.5.2.2 Paraglacial adjustment processes as system inherent controls	197
10.5.2.3 Topo-climatic forcing on permafrost rockwalls	200
10.6 Perspectives	202
10.7 Conclusion	203
11 Improving the performance of spatio-temporal models	206
11.1 Introduction	207
11.2 Case studies and description of the datasets	209
11.2.1 Case Study I: modelling air temperature in Antarctica	209
11.2.2 Case Study II: modelling volumetric water content of the "Cookfarm", USA	210
11.3 Methods	210
11.3.1 Random Forest algorithm	210
11.3.2 Validation strategies	212
11.3.3 Feature selection	212
11.4 Results and Discussions	214
11.4.1 Target-oriented validation	214
11.4.2 Detecting over-fitting	215
11.4.3 Reducing over-fitting and improving model performances	215
11.5 Conclusions	221
12 Conclusions	224
12.1 Significance of the developed products	224
12.2 Opportunities of machine learning in geography	226
12.3 Challenges of machine learning in geography	227
12.4 Accounting for space-time dependencies	228
12.5 Practical consequences	229
12.6 Outlook	231
12.7 Concluding remarks	231
References	233
Curriculum Vitae	274
Erklärung	279

List of Figures

1.1	A very simple interpolation of rainfall	6
1.2	A very brief description of the process of machine learning	9
1.3	Structure of this work	15
2.1	Flow chart of the main methodology applied in this study	21
2.2	Example of threshold tuning	25
2.3	Comparison of the rainfall area prediction performances of the four ML algorithms	31
2.4	IR image of the MSG SEVIRI scene from 2010/05/06 14:50 UTC	32
2.5	Visualization of the rainfall area predictions of the four ML algorithms for the exemplary scene	33
2.6	Comparison of the rainfall rate prediction performances of the four ML algorithms	34
2.7	Visualization of the 24-hour aggregated rainfall rate predictions of the four ML algorithms for the exemplary day	35
3.1	Overview of the methods to compare models that use spectral and textural variables with models that use spectral variables only	43
3.2	Dependence of the number of variables on the performance	47
3.3	Boxplots showing the performance of the full models as well as spectral-only models	48
4.1	Flow chart of the methodology applied in this study	54
4.2	Map of the average annual precipitation sums in the study area as estimated by WordClim	55
4.3	Validation of estimated rainfall areas for 2013 on an hourly basis	62
4.4	Comparison of POD for different hourly measured rainfall quantities as well as FAR for different predicted rainfall quantities	63
4.5	Validation of estimated rainfall quantities for 2013 on an hourly basis	63
4.6	Validation of estimated rainfall quantities for 2013 at (a) hourly resolution and on the different aggregation (b) daily, (c) weekly, (d) monthly	64
4.7	Monthly precipitation sums in mm of the year 2013 as estimated by this study	65
4.8	Comparison of the performance of the MSG based retrieval and GPM IMERG for rainfall area delineation between March and August 2014	66
4.9	Comparison of the performance of the MSG based retrieval and GPM IMERG for hourly rainfall quantities between March and August 2014	67
4.10	Sample satellite scene from 2014/04/24 10:00 UTC as well as the rainfall estimates for this scene, observed rainfall and GPM IMERG estimates	68

5.1	Map of weather stations used for model training and evaluation . .	76
5.2	Relative variable importance revealed by Cubist	80
5.3	Agreement between measured T_{air} and predicted T_{air} using Cubist as modeling tool	81
5.4	Correlation between measured and predicted T_{air} based on LOSOCV	82
5.5	Distribution of measured T_{air} compared to predicted T_{air}	82
5.6	Measured and predicted time series of three example weather stations	83
5.7	Monthly RMSE of T_{air} predictions for stations on ice and ice-free areas	84
5.8	Spatial distribution of RMSE of predicted T_{air} for the 32 weather stations	84
5.9	Performance of the T_{air} product on different aggregation levels . .	85
5.10	Monthly aggregates of T_{air} as predicted by the GBM model	85
6.1	Map of the study area including vegetation types as well as the position of the sampling locations	95
6.2	Schema of the processing flow	99
6.3	R^2 values of linear regressions between NBI and vegetation cover .	102
6.4	R^2 values of linear regressions between NBI and AGB	103
6.5	Example of the spatial predictions at the loation "Qumahe"	104
6.6	Validation of the vegetation cover models	105
6.7	Validation of the AGB models	105
6.8	Distribution of vegetation cover values at the different locations where satellite data was available	106
6.9	Distribution of biomass values at the different locations where satellite data was available	106
7.1	Overview of the processing flow	115
7.2	Map of the study area including the location of training images as well as images used for prediction	116
7.3	Variable importance of 15 highest ranked predictor variables in the Random Forests model	120
7.4	Reliability of 500 Google Earth images	122
7.5	Three example RGB images, the predicted probabilities for woody vegetation and the corresponding classification results	122
8.1	Cook Agronomy Farm overview map with soil profile sampling points and instrumented locations	132
8.2	Sensor values from five depths at one station at Cook Agronomy Farm	133
8.3	Distribution of observations for water content, temperature, and electrical conductivity across soil depth	134
8.4	Importance plots (covariates sorted by importance)	141

8.5	Spatio-temporal predictions of soil water content at Cook Agronomy Farm for the growing season in 2012 using the random forests (RF) model	142
8.6	Spatio-temporal sample variogram, metric variogram, and isolated 3D spatial and temporal components	144
8.7	Spatio-temporal predictions of soil water content at Cook Agronomy Farm for the growing season in 2012 using the kriging model .	145
8.8	Hexbin plots for observed and predicted values showing goodness of fit	146
8.9	Types of soil variables in terms of temporal stability or change . .	151
9.1	Comparison between measured and predicted soil respiration (SR) rates at 25°C and 45% of water holding capacity based on leave-one-out cross validation	164
9.2	Correlation between absorption at each wavenumber and SOC, $SR_{25,45}$, Q10, and SOC-degradability for the general dataset and the individual subsets	165
9.3	Baseline-corrected MIRS spectra of three grassland soils	165
9.4	Comparison between measured and predicted Q10 values based on leave-one-out cross validation	167
9.5	Comparison between measured and predicted soil respiration rates based on leave-one-sampling-point-out cross-validation	168
9.6	Comparison between measured and predicted soil respiration rates based on leave-one-sampling-point-out cross-validation for each land use type separately	169
9.7	Scaled variable importance as revealed by the Random Forest algorithm	169
10.1	Process-scale of potential rockfall controls	176
10.2	Study area	178
10.3	Selection of predictor variables	182
10.4	Modelling approach of the multiple logistic regression	184
10.5	Rockfall densities (RD in %) of predictor variables	187
10.6	Variable importance quantified by means of the random forest mode	191
10.7	Receiver Operating Characteristic	192
10.8	Typical examples for active talus slope deposition	197
10.9	Possible models for the timing of paraglacial rockfall activity in the Turtmann Valley	199
11.1	Schematic overview of validation strategies considered in this study	213
11.2	Differences in the Leave-Location-Out cross-validation performance of a) the air temperature estimations and b) the volumetric water content estimations using different feature selection strategies . . .	219
11.3	Relative scaled importance of the predictor variables within the Random Forest models for the case study of (a) T_{air} Antarctica and (b) VW Cookfarm	220

11.4 Differences in the Leave-Location-Out and Leave-Location-and-Time-Out cross-validation performance using no feature selection, a recursive feature elimination and the newly proposed forward feature selection	220
--	-----

List of Tables

2.1	Summary of rainfall areas and rainfall rates of the three input datasets	23
2.2	Overview of the predictor variables used to model rainfall areas and rainfall rates	24
2.3	Optimal tuning parameters	27
2.4	Confusion matrix as baseline for the calculation of the verification scores	28
2.5	Calculation of the confusion matrix-based verification scores	28
2.6	Processing time in minutes for model tuning and training as well as just model training with the optimal tuning parameters	36
4.1	Optimal hyperparameters for the individual models	60
4.2	Confusion matrix as baseline for the calculation of the verification scores	61
4.3	Categorical metrics for validation of rainfall area estimates	61
5.1	Tested values for the hyperparameters of the different prediction models and the optimal parameters revealed during parameter tuning	79
6.1	Overview of the 18 locations which were sampled during field work in 2011 and 2012	96
6.2	Summary statistics of the vegetation cover and AGB samples	97
6.3	Overview of the spectral bands of the three satellite sensors Quick-Bird, RapidEye and WorldView-2	98
6.4	Summary of the satellite images which were available for this study	98
6.5	Cross validated results of the feature selection with RF training for the different models calculated in this study	102
7.1	Confusion matrix	118
7.2	Cross tabulation based validation metrics	119
7.3	Contingency table of the test data set	120
7.4	Performance of the Random Forests model	120
7.5	Excerpt of the training site database for the three example images shown in Fig. 7.5.	121
8.1	Cook Agronomy Farm data set spatio-temporal covariates	135
8.2	Parameters of the seasonality functions for water content (VW), soil temperature (C) and electrical conductivity (EC)	143
8.3	Variogram parameters for each variable	143
8.4	Global cross-validation statistics	147
9.1	Environmental soil classes (ESC) and sample set	158
9.2	Prediction accuracy for $SR_{25,45}$, Q10 values, and C-degradability (i.e., $SR_{25,45}/SOC$, SOC normalized soil respiration) based on the different submodels	166
10.1	Overview on response and predictor variables	180

10.2	Diagnostics statistics of multicollinearity between independent predictors	185
10.3	Varimax-rotated principal components of original standardised predictor variables	189
10.4	Test for goodness-of-fit	190
10.5	Coefficient statistics	190
10.6	Contingency table for the principal component logistic regression model and the random forest model	192
11.1	Predictor variables used within the two case studies	211
11.2	Regression statistics between observed and predicted values based on cross-validation	216

List of acronyms

AGB	Aboveground Biomass
AIC	Akaike Information Criterion
AUC	Area Under the Curve
AVNNET	Averaged Neural Network
BLD	Bulk density of soil
Bt	Occurrence of Bt horizon
cdayt	Transformed cumulative day
CLM	Community Land Model
CSI	Critical Success Index
CV	Cross-Validation
DEM	Digital Elevation Model
DTM	Digital Terrain Model
EC	Electrical Conductivity
ESC	Environmental Soil Classes
ETS	Equitable Threat Score
FAR	False Alarm Ratio
FFS	Forward Feature Selection
GLCM	Grey Level Co-occurrence Matrix
GBM	Generalized Boosted Regression Models
GPM	Global Precipitation Measurement
HKD	Hansen-Kuipers Discriminant
HSS	Heidke Skill Score
IMERG	Integrated Multi-satellite Retrievals for GPM
IR	Infrared
LGM	Late-Glacial Maximum

LLO	Leave-Location-Out
LLTO	Leave-Location-and-Time-Out
LOSO CV	Leave-One-Station-Out Cross-Validation
LR	Logistic Regression
LST	Land Surface Temperature
LTO	Leave-Time-Out
MAE	Mean Absolute Error
ME	Mean Error
MIRS	Mid-Infrared Spectroscopy
ML	Machine Learning
MODIS	Moderate-resolution Imaging Spectroradiometer
MSG	Meteosat Second Generation
mtry	Number of variables randomly sampled at each split
NBI	Narrow-Band Indices
NDRE	Normalized Difference Red-Edge Index
NDVI	Normalized Difference Vegetation Index
NIR	Near Infrared
NNET	Neural Network
ntree	Number of trees
PC	Principal Components
PCLR	Principal Component Logistic Regression
PHI	Soil pH
PLSR	Partial Least Square Regression
POD	Probability Of Detection
POFD	Probability Of False Detection
Q10	Increase of soil respiration by a 10°C increase of temperature

QB	Quickbird
QTP	Qinghai-Tibet-Plateau
R²	Coefficient of determination
RAMP	Radarsat Antarctic Mapping Project
RE	Rapid Eye
RF	Random Forest
RFE	Recursive Feature Elimination
RMSE	Root Mean Square Error
ROC	Receiver Operating Characteristic
sd	Standard deviation
SEVIRI	Spinning Enhanced Visible and Infrared Imager
SOC	Soil Organic Carbon
SR	Soil Respiration
SR_{REF}	Soil Respiration at a reference temperature
SVM	Support Vector Machines
T_{air}	Air Temperature
T_{REF}	Reference Temperature
TRMM	Tropical Rainfall Measuring Mission
UTC	Coordinated Universal Time
VIF	Variation Inflation Factors
VIS	Visible
VVI	Visible Vegetation Index
VW	Volumetric Water Content
WHC	Water Holding Capacity
WV	Water Vapour

Chapter 1

Introduction

1 Introduction

One of the key tasks in environmental geography is obtaining information of geographic features in space or in space and time. Considering current trends towards *big data*, increasing volume, velocity, and variety of geographic data (van Zyl, 2014) lead to new opportunities for environmental monitoring that are accompanied with a paradigm shift towards data-driven data analysis. In this context, machine learning algorithms learn patterns in nonlinear and complex systems. That makes them an important tool in environmental geography that is highly associated with nonlinearity and complex underlying interactions. Whilst the number of machine learning applications in environmental geography rapidly increases, the characteristics of geographic data (especially spatial dependencies) remain widely unconsidered in geographic machine learning applications.

The following introduction will give an overview on the modelling tasks in environmental geography and the necessary paradigm shift towards data-driven model development. Based on limitations and challenges of recent machine learning applications, the aim and hypotheses of this thesis are developed followed by a description of the general outline of this thesis.

1.1 Modeling tasks in environmental geography

To understand the potential for data-driven modelling in environmental geography, it is worth clarifying the major tasks in geography that require modelling approaches.

- Mapping of geographic features

A frequent task in geography is obtaining spatially explicit information of environmental features based on limited field observations. Thus, small scale data records are transferred into space to obtain maps of the feature of interest. Mapping of geographic features is a common task in all fields of environmental geography. In biogeography, mapping of land cover (Gómez *et al.*, 2016) or biodiversity (Miller, 1994) are common applications. In soil science and geomorphology, mapping of soil characteristics (McBratney *et al.*, 2003; Brevik *et al.*, 2016) and geomorphological features (Smith *et al.*, 2011) are routine. The aims of spatial mapping are diverse and serve several purposes like policy making and planning of e.g. conservation areas according to biodiversity characteristics (Ferrier, 2002). Spatial mapping in geography is further used as a tool for risk assessment, e.g. of geomorphological hazards (Reichenbach *et al.*, 2005; Lee and Sambath, 2006) or

flooding (Porter and Demeritt, 2012). In addition, spatially explicit data serve as essential baseline products that subsequent scientific studies can build upon.

- Spatio-temporal monitoring of geographic features

Whilst some geographic features can be considered as being temporally comparably static (e.g. soil types), other features are highly dynamic not only in space but also in time (e.g. soil moisture). Therefore, spatio-temporal monitoring extends the approach of temporally static mapping of geographic features by considering temporal dynamics. The aim of a spatio-temporal monitoring is to obtain dynamics of a certain feature in space and time. Potential areas for application in environmental geography are the monitoring of dynamic vegetation characteristics as e.g. phenology (Zhang *et al.*, 2003). In climatology, most variables of interest are even more dynamic than vegetation characteristics, as for example air temperature (Hengl *et al.*, 2011) or rainfall (Kidd and Huffman, 2011). The variability of climate has an impact on other spheres that react in a highly dynamic way. Soil temperature or soil moisture for example react on the dynamic climatic impacts and its spatio-temporal monitoring is an important field in soil science (Gasch *et al.*, 2015). Spatio-temporal monitoring allows analysing dynamics and trends and form valuable tools for planning, policy making (Ceccato *et al.*, 2014) and as baseline products for scientific studies.

- Forecasting

Spatio-temporal dynamics are not only studied from past and present times, but predictions are also made into the future. Forecasting is especially relevant in the field of climatology with regard to short term weather forecasts and long-term trends. Climate change predictions are the most prominent example in this context (IPCC, 2014). Forecasting also plays a role in other disciplines, for example concerning projections of land use and land cover change (Veldkamp and Lambin, 2001; Thies *et al.*, 2014). By building scenarios of future behaviour, forecasting is an essential tool for policy making and to develop strategies for adaptation (IPCC, 2014).

- Enhancing knowledge about system behaviour

Another aim in environmental geography that requires modelling is the understanding of system behaviour. It is a question of how individual components influence a system and how a system reacts to changing conditions (Bossel, 1994). Examples from the field of biogeography and climatology could be gaining knowledge about vegetation-atmosphere interactions (Krinner *et al.*, 2005) while in geomorphology the delineation of factors influencing the risk of landslides are relevant (Vorpahl *et al.*, 2012).

As it is apparent from the list of modelling tasks, there are two general targets pursued: Mapping, monitoring as well as forecasting of geographic features aim at creating accurate maps, time series or scenarios, while the second target is associated with system understanding and the identification of driving forces that lead to those spatial or spatio-temporal patterns. These two general targets are approached with two different categories of models: predictive models and explanatory models (Shmueli, 2010).

Predictive models (Kuhn and Johnson, 2013a) in environmental geography are mainly statistical. Such models are built upon the statistical relationship between field data of the target and spatially available predictor data (e.g. remote sensing data). As a very simple example, we could assume the task of mapping air temperature using the assumption of decreasing air temperature with increasing elevation. The statistical model is established from ground observations of air temperature (i.e. via climate stations) and corresponding information about elevation. The developed model can then be applied on the entire set of the spatially available data (i.e. digital elevation model) to obtain spatially explicit temperature estimates. While predictive models aim at accurate estimations of a feature in space and time (i.e. monitoring of air temperature), the explanatory models aim at an accurate understanding of processes and interactions (i.e. how is elevation related to air temperature?). Explanatory models can be statistical where potential influencing factors are tested for their relationship to the target, or conceptual or physical (Gray and Gray, 2017) where the model aims at representing a simplified representation of a system.

In this thesis, I will focus on predictive models, however, with consideration of explanatory components. Main emphasis will be on modelling strategies for spatial mapping and spatio-temporal monitoring. Though most of the content is applicable to the modelling task of forecasting as well, forecasting is not explicitly considered in this thesis.

1.2 Predictive modelling strategies

In view to the task of spatial mapping and spatio-temporal monitoring, it is a question of how the spatial or spatio-temporal dynamics of a parameter can be obtained. The initial situation is that we usually have point data (e.g. from climate stations) or data from small scale surveys (e.g. vegetation plot records) available that give us the geographic feature of interest for certain spatial locations and at certain points in time. However, initially we do not know anything about the feature's behaviour beyond the sample locations and beyond the date of the survey.

In the following sections, the pathway from point or small scale data to spatially explicit and temporally continuous data will be outlined. First, the "classic" approaches will be discussed followed by the delineation of the need towards data-driven model development.

1.2.1 Classic modelling strategies

1.2.1.1 Spatial Interpolations

The most obvious approach to obtain spatially explicit data might evolve from Tobler's first law of geography that implies "*everything is related to everything else, but near things are more related than distant things*" (Tobler, 1970). This law established the basis for the concept of spatial interpolations. The principle of spatial interpolations is that the characteristics of a feature are spatially correlated and therefore, point data can be transferred into space according to the distance.

Let's consider the task of spatio-temporal monitoring of hourly rainfall to illustrate the idea and problems associated with the concept of spatial interpolations. As an example, we have a number of climate stations distributed over southern Africa, that measure precipitation on an hourly resolution. However, we want to know rainfall for the entire area of southern Africa. According to the idea of interpolation, recorded rainfall values at a certain point in time (Fig. 1.1A) are interpolated by considering the distance to the climate stations (e.g. using a simple kriging interpolation, Fig. 1.1B). However, two problems are associated with this approach. The first problem becomes obvious by visual interpretation of the resulting map (Fig. 1.1B). Though the interpolation might have produced reliable results in the areas where the density of climate stations was high, the results become highly unreliable in areas with a low density of stations. Obviously, as rainfall is a highly dynamic parameter, a spatial interpolation that simply bases on the distance to the weather stations does not produce satisfying results. Admittedly, the example shown here is very simple. Though it could be extended by using a more complex interpolation approach, e.g. including further explanatory predictors as e.g. elevation (Goovaerts, 2000) or by using more complex algorithms (Ly *et al.*, 2011), there is a second problem that is associated with this concept. Using spatial interpolations, monitoring is restricted to the date where field records are available as the concept relies on the field data. Thus, if no field data are available, no spatial mapping can be performed. Considering the example of rainfall monitoring, this problem might be of low relevance since modern data loggers quasi-continuously produce data. However, other studies rely on temporally restricted data when field surveys are of high temporal and economic costs (e.g. vegetation surveys).

Due to the continuous dependence on field data as well as the lack of a suitability for dynamic variables, spatial interpolations do not provide a comprehensive and satisfying solution for the task of spatio-temporal monitoring when dealing with highly dynamic features and spatially or temporally limited ground truth data. Therefore, other approaches are required.

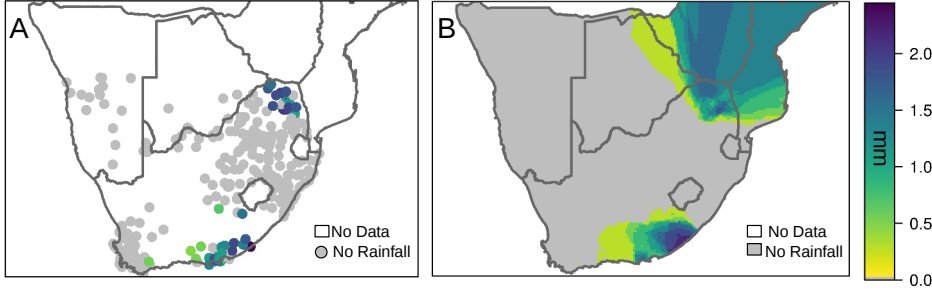


Figure 1.1: A very simple interpolation of rainfall in southern Africa from 2014/04/24 10:00. A shows the measured rainfall from several climate stations (Meyer *et al.*, 2017a). B shows the results from a simple kriging approach.

1.2.1.2 Statistical parametric models

Another well-established way for spatial mapping or spatio-temporal monitoring uses spatially available proxies or predictors for the feature of interest. With regard to remote sensing, there is much information available from space, that can be related to geographical features or processes by regression or classification analysis. Examples include the increase of biomass with increasing satellite-retrieved NDVI (Gizachew *et al.*, 2016), the relationship between satellite-based surface temperature and air temperature (Vogt *et al.*, 1997), or the increasing probability for rainfall with decreasing cloud temperatures that as well are provided by satellites (Vicente *et al.*, 1998).

Once a statistical model is built between satellite data and the response variable, it can be applied to the full extent of the satellite scene, or even to a time series, allowing spatial mapping or spatio-temporal monitoring of the response (e.g. Gizachew *et al.*, 2016; Shi *et al.*, 2016; Lopresti *et al.*, 2015, to mention just a few). By using this strategy, we got rid of the dependency on continuously available field data because once the model is built, no further ground truth data are required.

A model using just one predictor variable might, however, only in rare cases provide a good estimate of the response variable. Usually more than a single predictor is required to explain a feature's characteristic. Though common parametric models can also be of a more complex form and include more than one predictor (e.g. Lakshmi, 2013; Badreldin and Sanchez-Azofeifa, 2015; Lin *et al.*,

2012), the parametric approaches have a significant limitation: they are based on an *a priori* assumption of the data distribution (Breiman, 2001b) as well as of the form of the relationship between predictors and response (James *et al.*, 2013). This form is often assumed to be linear but can also be exponential or even more complex. While it is still possible to assess the appropriate relationship between predictor and response when only one or very few predictors are used, it becomes nearly impossible to assess the individual relationships when a large number of predictors is considered. A large number of predictors is further problematic in view to multicollinearity (Graham, 2003). Since most geographic data are correlated (e.g. spectral reflectance in different wavelengths), and this behaviour cannot be included in most parametric models, a considerable reduction in the number of predictors is often necessary from a technical perspective. Most systems, however, can only be described by a large number of interacting predictor variables and behave in a nonlinear manner. This problem raises the need for more flexible models that can handle large numbers of predictors, different types of variables and arbitrary relationships.

1.2.2 The paradigm shift towards data-driven model development

As more and more (spatial) data become available the requirement for more flexible models force a paradigm shift towards true data-driven model development (Miller and Goodchild, 2015). Data-driven model development, or what Breiman (2001b) refers to as *algorithmic modelling* in contrast to *data modelling*, is a designation associated with *big data* and aims at finding relationships in the data without an *a priori* assumption about the system (Breiman, 2001b; Lary *et al.*, 2016). In this context, machine learning algorithms are applied as a predictive modelling tool to learn arbitrary relationships in the data and to make predictions according to the learned function. The advantage compared to parametric approaches is that machine learning algorithms learn relationships between predictors and responses by themselves. This allows a greater flexibility and the utilization of many, correlated, or even potentially uninformative predictor variables. In this context it is of note that the greater flexibility is at the expense of interpretability. Machine learning algorithms are referred to as a *black box* because the exact learned relationship between predictors and response is difficult to interpret (Lary *et al.*, 2016). However, machine learning algorithms are advantageous when prediction is in the foreground rather than an exact understanding of underlying relationships. Therefore, they have high potential for mapping or monitoring of geographic features.

In general, we can distinguish between two categories of learning: supervised learning and unsupervised learning (James *et al.*, 2013). Supervised learning is based on *training data* that consist of predictor variables and a response (i.e.

measurements). The aim of supervised applications is to learn how the predictors can best describe the response. In contrast, unsupervised learning is based on predictor variables solely, thus there is no response variable and the data are considered to be *unlabelled*. The aim of unsupervised learning is then to structure the predictor variables in a way that a subsequent labelling of the data is possible. According to the major modelling tasks outlined in section 1.1, this thesis focus on supervised learning tasks. The fundamentals of supervised machine learning will be explained in the following.

1.3 Machine learning fundamentals and terminology

Machine learning is a collective term for a variety of data-driven algorithms that aim at learning the relationship between predictor and response variables and make predictions based on the developed model. A usual modelling task (Fig. 1.2) starts with the acquisition of the target variable which is also referred to as *response variable*, *dependent variable* or *ground truth*. Usually this is data taken from field surveys or data loggers. *Predictor variables* (also referred to as *independent variables*) are then required to estimate the *response variable*.

Predictor and response variables form the initial dataset which is then split into *training* and *testing* data, thus into data that is used to *train* the model, and data that is used to *validate* or *test* the model. Based on the training data, a machine learning algorithm then learns the relationship between predictors and response which is designated as *model training*. Most machine learning algorithms have so called *hyperparameters* or *tuning parameters* that control the model complexity which cannot be directly estimated from the data (Kuhn and Johnson, 2013a). Therefore, a *tuning* of these parameters must be included in the process of model training.

Both, model tuning and model training, must always be evaluated in view to *independent data*, thus the effect of the hyperparameters as well as the performance of the final model must always be evaluated with held back data. If this is not considered, the resulting model has a high risk of over-fitting because highly complex models are able to fit to noise in the training data. This, however, is not applicable for the general relationship between predictors and response (James *et al.*, 2013). In this context, cross-validation has been evaluated as a robust tool to tune the complexity of models to avoid over-fitting (Kuhn and Johnson, 2013a; James *et al.*, 2013). For cross-validation, the data are split into several *folds* (*resamples*). For each iteration, a model is trained with a respective set of hyperparameters using all data except for one fold and the performance of the model to predict on the held back fold is then assessed. In this way, the optimal set of hyperparameters can be retrieved and an objective performance metric of

the final model can be given.

Once the model is trained, it can be applied on an entire set of predictors to make *predictions* in space, or on a new set of predictors to make predictions in space and time. It is of note, that the term *prediction* in this context is not synonymous to the term *forecast*. Though predictions might be made for future conditions, the term more generally means to estimate the response variable for unknown data, thus for locations where no ground truth data were available or for unknown points in time (within a defined model domain).

Finally, it is of note that supervised machine learning can aim at two different tasks: *regression* or *classification*. While the response of classification models is *categorical* (e.g. land cover classes), the response of regression models is *continuous* or *numeric*.

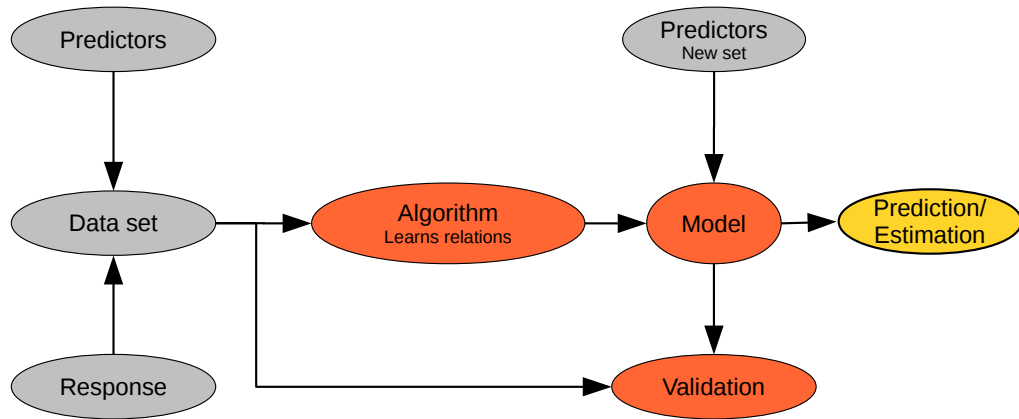


Figure 1.2: A very brief description of the process of machine learning. The grey colored shapes represent data, orange the modelling procedure and yellow the outcome.

1.4 Machine learning in environmental geography - State of the art

Machine learning algorithms are well-established in environmental sciences (Lary *et al.*, 2016; Kanevski *et al.*, 2009; Hsieh, 2009) and find application in all fields of environmental geography. In this context, machine learning is widely used in conjunction with remote sensing (see review in Paola and Schowengerdt, 1995; Lary *et al.*, 2016; Camps-Valls, 2009; Mountrakis *et al.*, 2011) as it provides an excellent source for spatial and spatio-temporal predictor variables for a variety of environmental research tasks. The following section gives a brief - by no means exhaustive - overview where machine learning is used in different fields of environmental geography.

1.4.1 Biogeography

One of the typical applications of machine learning in the field of biogeography is the mapping of land use/cover based on optical satellite information (Gislason *et al.*, 2006; Rodriguez-Galiano *et al.*, 2012). In this context, machine learning algorithms, as for example neural networks or support vector machines, have shown to be superior compared to traditional methods such as the maximum likelihood classifier (Huang *et al.*, 2002; Otukei and Blaschke, 2010; Waske *et al.*, 2009). Using machine learning and multispectral data, land cover could be classified into broad vegetation types and the use of hyperspectral data allowed further classification down to a species level (Baldeck *et al.*, 2015; Lawrence *et al.*, 2006). Multispectral, as well as hyperspectral data, in conjunction with machine learning are further used to map vegetation cover (Lehnert *et al.*, 2015b), biophysical characteristics (Verrelst *et al.*, 2012), biomass (Ali *et al.*, 2015) or tree diversity (Vaglio Laurin *et al.*, 2014). Ground truth data for these studies were usually provided by field surveys where vegetation characteristics were sampled on a plot scale.

Whilst spectral satellite data can be considered to be directly related to vegetation patterns, machine learning was used in modelling tasks, where more indirect predictor variables were applied. Baltensperger and Huettmann (2015) modelled the diversity of mammals in Alaska using derived remote sensing products that included land cover, climatological information as well as terrain properties. Habitat suitability was also modelled with machine learning on derived remote sensing products, for example to obtain potential habitats for *Pinus sylvestris* on the Iberian Peninsula (Garzón *et al.*, 2006).

1.4.2 Climatology

Machine learning has a long-term history in the field of spatial atmospheric science. Cloud type classifications (Tian *et al.*, 1999; Giannakos and Feidas, 2013), cloud characteristics (Jung *et al.*, 1998) as well as rainfall (Hsu *et al.*, 1997; Hong *et al.*, 2004; Behrangi *et al.*, 2009a; Kühnlein *et al.*, 2014b) were modelled using machine learning. With a few exceptions (Kühnlein *et al.*, 2014a,b), artificial neural networks are the prevailing algorithms in the field of cloud and rainfall modelling. This is in contrast to other fields of environmental geography, where a high variability of algorithms is applied. The idea behind cloud and rainfall retrievals is that the spectral information (e.g. optical, Kühnlein *et al.*, 2014b) is related to cloud properties which are further related to rainfall probabilities. As well as monitoring climatic patterns, machine learning was applied as an alternative statistical downscaling approach for general circulation models (Tripathi *et al.*, 2006).

As climate provides boundary conditions for other systems (Bonan, 2008), climate monitoring products are of high relevance for subsequent studies, e.g. as important predictors for biodiversity mapping (Baltensperger and Huettmann, 2015).

1.4.3 Soil science and hydrology

The application of machine learning in soil science and hydrology is rather recent but of increasing interest to the scientific community. A large field of application is mapping of soil taxonomic units (see review in Heung *et al.*, 2016) but also mapping of soil properties like soil moisture (Ahmad *et al.*, 2010; Ali *et al.*, 2015), soil organic carbon (Ließ *et al.*, 2016; Henderson *et al.*, 2005), or nitrogen and phosphorus content (Henderson *et al.*, 2005). As ground truths, point observations from soil samples or soil profiles are being used and the response variable is usually predicted from topographic information, spectral satellite data and/or climate indices (Heung *et al.*, 2016).

Another application of machine learning that also has the aim to provide high resolution soil moisture datasets is the downscaling of low resolution satellite-based soil moisture products with higher resolution predictors (Srivastava *et al.*, 2013; Im *et al.*, 2016).

In regard to hydrology, machine learning finds frequent application in stream-flow modelling and forecasting (Rasouli *et al.*, 2012; Asefa *et al.*, 2006; Shortridge *et al.*, 2016). However, these applications are usually not spatially explicit but focus on temporal patterns. Space as well as time, however, recently found consideration in machine-learning based run-off modelling (Gudmundsson and Seneviratne, 2015). Further hydrological applications of machine learning are compiled in Govindaraju and Rao (2000).

1.4.4 Geomorphology

The most common application of machine learning in geomorphology is the mapping of landslide susceptibility (Micheletti *et al.*, 2014; Catani *et al.*, 2013; Goetz *et al.*, 2015; Brenning, 2005) which reaches into the field of risk assessment. Machine learning was also applied to map landforms, for example types of glaciated landscapes (Brown *et al.*, 1998). Machine learning applications for geomorphology are further reviewed in Valentine and Kalnins (2016).

1.5 Formulation of the scientific problem, aims and hypotheses

As outlined in section 1.4, machine learning is used in all fields of environmental geography and the number of applications is considerably increasing. Machine learning, however, is not a very recent discovery in environmental geography. In contrast, machine learning to obtain spatio-temporal datasets from limited ground truth data was already applied in the 1990s, however, at this time, due to the high complexity of application, it was only used by experts in this field. Major developments in software packages in recent years, allow greater access to machine learning for virtually everyone. Well-known GIS software (ArcGIS, SAGA, QGIS, GRASS, IDRISI, etc) provides easily accessible machine learning functionality for environmental mapping. Especially R, as a frequently used software in natural science, has a variety of machine learning algorithms implemented (Hothorn, 2017). The caret package for R (**C**lassification **A**nd **R**egression **T**raining, Kuhn, 2016a) allows access to most of the implemented algorithms via a handy and unified syntax and further provides a variety of functions for data processing, model tuning and evaluation, parallel computing as well as model visualization.

While the large number of machine learning applications in scientific journals, as well as the today user-friendly software, feign a simplicity of machine learning, the complexity of the methods has not changed with time. Underestimating the technical complexity increases the risk of incorrect utilization of algorithms and can lead to false interpretations and conclusions. This is especially problematic in the field of geography since machine learning algorithms were not originally developed for spatial and spatio-temporal data analysis and the common workbooks that serve as guidelines on how machine learning is used (e.g. Kuhn and Johnson, 2013a; James *et al.*, 2013) do not refer to geographic data. Therefore, machine learning algorithms in geography are usually applied in the same way as in statistical medicine, economics, and other non-spatial fields. However, geographic data have certain characteristics (see section 1.5.1) that make them stand out against the mass of "ordinary" data and this should have serious consequences for the utilization of machine learning in geography.

1.5.1 Characteristics of geographic data

The most obvious and important characteristic of geographic data is surely its localisation in space (*geospatial data*). Geospatial data refer to a location on ground and provide data of a variable at the corresponding location. Vector point data might be the most intuitive example for geospatial data and the most frequent type of ground truth data for prediction models. Point data can be linked to a point on earth by its coordinates and a reference system and include certain information about a geographic feature at this point (e.g. via data loggers

on climate stations). In contrast, the majority of earth observation data and the main source of predictor variables being used in predictive models are provided by remote sensing (Lary *et al.*, 2016) and typically represented as raster data. Raster data provide discrete or continuous information of geographic features in a spatially explicit way. Especially such spatially explicit data illustrate one of the key characteristics of geospatial data: they are within a certain degree dependent in space causing spatial patterns to mostly appear as either patches or gradients (Legendre, 1993). This dependence of a variable in space causes *spatial autocorrelation*. Spatial autocorrelation means that samples are more/less similar than what might be expected from a random distribution (Legendre, 1993). The geographic feature at location "x" depends in a certain way on the feature on neighbouring locations and/or on the environmental characteristics not only of the location "x" itself but also of the neighbourhood. Thus, observations (i.e. spatial pixels or points) are not independent of each other.

Autocorrelation of geospatial data does not only happen in space but also in time (Shen *et al.*, 2016). Especially when time series of geographic features are to be analysed (e.g. air temperature, soil moisture, vegetation greenness), the observations at one point often feature a *temporal autocorrelation* resulting in dependency in space and time (*spatio-temporal autocorrelation*).

Though spatial and spatio-temporal autocorrelation is probably the key characteristic, geographic data have further characteristics that might be important in view to machine learning applications. The irregularity of many geographic features lead to unbalanced data, e.g. considering the ratio between raining and non raining clouds (Kühnlein *et al.*, 2014b) which is highly unbalanced as the averaged proportion of non raining clouds is considerably higher compared to raining clouds. Further, geographic datasets feature a large variability in size. Often large amounts of potential predictor variables (e.g. via remote sensing) contrast with only a few samples of response variables (e.g. vegetation surveys).

These characteristics (especially the spatial and/or temporal dependency) make geographic data special when compared to the standard data used in other scientific fields. This raises the question if, and how, these characteristics need to be incorporated in machine learning applications.

1.5.2 Aims and hypotheses

This thesis aims at assessing the potential and sensitivity of machine learning for environmental geography. In this context, as the characteristics of geographic data are being widely ignored in the large amount of recent machine learning applications, the superior aim of this thesis is to advance the field of machine learning in environmental geography by addressing these characteristics. The

thesis is developed in view to the hypothesis that, only if the characteristics of geographic data are considered, data-driven modelling strategies lead to a gain of information and to robust spatio-temporal model results.

Therefore, based on a variety of environmental monitoring applications, the thesis aims at developing adequate modelling strategies with respect to characteristics of geographic data, to provide reliable spatio-temporal data from limited field observations that support knowledge about different ecosystem components. The series of applications provides the basis to discuss the influence of geographic data and consequent modelling strategies in the general context of machine learning-based spatio-temporal monitoring of the environment.

1.5.3 Concept and structure of this thesis

During this thesis, a number of contributions in a broad spectrum of environmental geography have been published so that this thesis presents a collection of comprehensive knowledge about machine learning in environmental geography. The individual contributions are incorporated in the major hypothesis that, if characteristics of geospatial data are not considered, data-driven modelling strategies lead neither to a gain of information nor to robust spatio-temporal model results. Each chapter further provides new insights or monitoring products in its respective field of research.

Within this thesis the individual contributions are structured according to their field of research (Fig. 1.3). Chapters 2, 3 and 4 thematically focus on rainfall retrieval based on optical satellite data. Chapter 5 further covers this climatological context and aims at developing a spatio-temporal satellite-based monitoring product of air temperature for Antarctica. From the field of biogeography, chapter 6 evaluates different hyperspectral and multispectral indices to map vegetation cover and biomass on the Qinghai-Tibet-Plateau. Chapter 7 presents a method to automatically create Google Earth based training data for a larger scale monitoring of bush encroachment in South Africa. From the field of soil science, chapter 8 addresses modelling soil properties in space, time and depth on a farm scale and chapter 9 aims at developing a model to obtain soil respiration estimates from mid-infrared data. In a geomorphological context, chapter 10 aims at identifying factors that lead to rockfall in the Turtmann Valley in the Swiss Alps. The final publication in this thesis (chapter 11) wraps up the findings from the individual case studies by addressing the problem of spatio-temporal over-fitting due to the characteristics of geospatial data.

The key findings of this thesis could only be a result of a development process and are consequently maturing over the individual publications. Therefore, the final chapter of this thesis (chapter 12) summarizes the major methodological

developments from this study and discusses them in the broader methodological context. This chapter will further highlight the scientific outcome of the individual chapters and give recommendations for the utilization of machine learning in environmental geography.

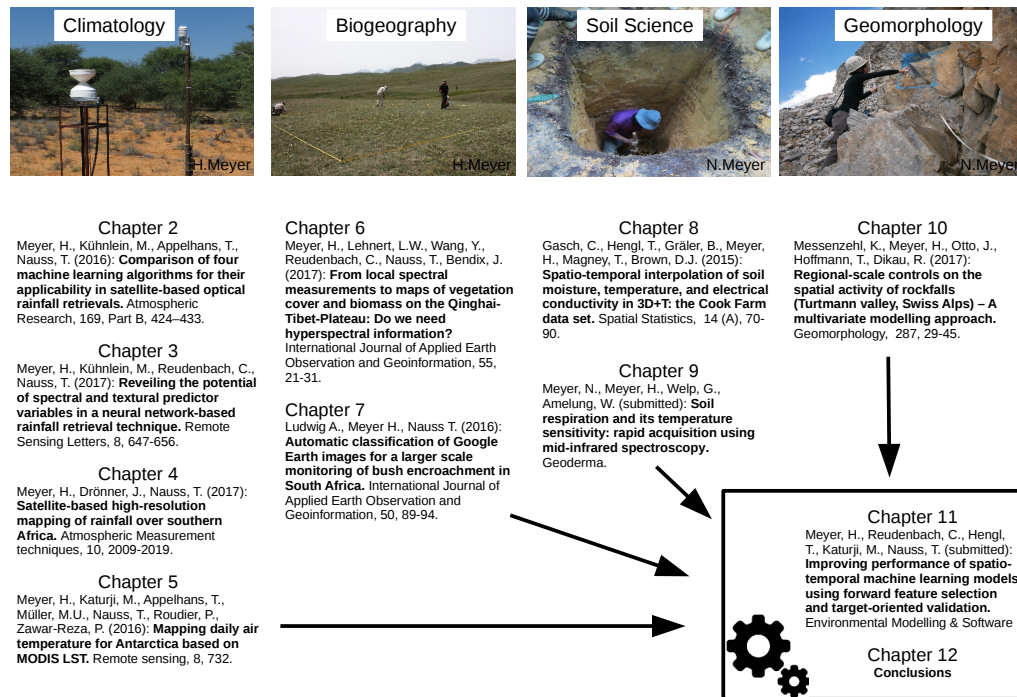


Figure 1.3: *Structure of this work*

Chapter 2

Comparison of four machine learning algorithms for their applicability in satellite-based optical rainfall retrievals

Hanna Meyer (1), Meike Kühnlein (1), Tim Appelhans (1),
Thomas Nauß (1)

- (1) Environmental Informatics, Faculty of Geography, Philipps-University Marburg, Deutschhausstr. 10, 35037 Marburg, Germany

Published in

Atmospheric Research 2016, 169 Part B, 424–433

Received 10 December 2014

Revised 19 September 2015

Accepted 22 September 2015

Available online 8 October 2015

Published 1 March 2016

2 Comparison of four machine learning algorithms for their applicability in satellite-based optical rainfall retrievals

Hanna Meyer, Meike Kühnlein, Tim Appelhans, Thomas Nauß

Abstract

Machine learning (ML) algorithms have successfully been demonstrated to be valuable tools in satellite-based rainfall retrievals which shows the practicability of using ML algorithms when faced with high dimensional and complex data. Moreover, recent developments in parallel computing with ML present new possibilities for training and prediction speed and therefore makes their usage in real-time systems feasible. This study compares four ML algorithms - random forests (RF), neural networks (NNET), averaged neural networks (AVNNET) and support vector machines (SVM) - for rainfall area detection and rainfall rate assignment using MSG SEVIRI data over Germany. Satellite-based proxies for cloud top height, cloud top temperature, cloud phase and cloud water path serve as predictor variables. The results indicate an overestimation of rainfall area delineation regardless of the ML algorithm (averaged bias = 1.8) but a high probability of detection ranging from 81% (SVM) to 85% (NNET). On a 24-hour basis, the performance of the rainfall rate assignment yielded R^2 values between 0.39 (SVM) and 0.44 (AVNNET). Though the differences in the algorithms' performance were rather small, NNET and AVNNET were identified as the most suitable algorithms. On average, they demonstrated the best performance in rainfall area delineation as well as in rainfall rate assignment. NNET's computational speed is an additional advantage in work with large datasets such as in remote sensing based rainfall retrievals. However, since no single algorithm performed considerably better than the others we conclude that further research in providing suitable predictors for rainfall is of greater necessity than an optimization through the choice of the ML algorithm.

Keywords Machine learning; Rainfall retrieval; Rainfall rate; Rainfall area; MSG SEVIRI

2.1 Introduction

Spatially explicit, continuous and high-resolution monitoring of precipitation is important for a variety of fields in the environmental sciences as well as for the economy and society as a whole. Satellite-based methods are currently the only way to fulfill the requirement of area-wide information. Amongst the variety of available satellite systems, optical sensors on-board geostationary satellites offer high spatial and temporal resolution, which is important when considering local and short-term rainfall events (Thies and Bendix, 2011). Furthermore, the latest systems feature adequate spectral resolutions for detecting cloud-top properties such as cloud top height, cloud top temperature, cloud phase and cloud water path (Thies *et al.*, 2008b).

Over the last several decades, many optical satellite-based rainfall retrieval techniques for the detection of precipitating clouds and assignment of rainfall rates have been developed (see valuable overviews by Kidd and Levizzani, 2011; Prigent, 2010; Thies and Bendix, 2011; Kidd and Huffman, 2011; Levizzani *et al.*, 2002). These retrievals are generally based on parametric relations between spectral properties as proxies for cloud-top properties, rainfall areas and rainfall rates. Rainfall areas are commonly delineated from non-raining clouds using thresholds in selected satellite channels and/or derived information (Ba and Gruber, 2001; Feidas and Giannakos, 2012; Roebeling and Holleman, 2009; Thies *et al.*, 2008b,a). Rainfall rates are then assigned by relating the spectral information to measured or modelled rainfall rates (Adler and Negri, 1988; Kühnlein *et al.*, 2010; Roebeling and Holleman, 2009; Vicente *et al.*, 1998).

The parametric techniques used within rainfall retrievals have the advantage that they directly map the conceptual knowledge of rainfall processes to their retrieval using remotely sensed proxies. However, machine learning (ML) approaches have generally been shown to be superior when the prediction, and not the understanding of underlying processes, is the focus (Kuhn and Johnson, 2013a). Moreover, parametric approaches usually consider only a limited number of predictor variables while ML algorithms can handle the full set of available information.

Precipitation processes leading to different rainfall intensities are very complex. In this context ML algorithms have been deemed valuable tools for dealing with complexity, non-linearity and highly correlated predictor variables. Neural network algorithms are most frequently used in rainfall retrieval techniques to link the input information to rainfall estimates (Behrangi *et al.*, 2009b; Capacci and Conway, 2005; Giannakos and Feidas, 2013; Grimes *et al.*, 2003; Hong *et al.*, 2004; Hsu *et al.*, 1997; Rivolta *et al.*, 2006; Tapiador *et al.*, 2004). Random forests is an ensemble technique commonly applied in remote sensing especially for land

cover classifications (Gislason *et al.*, 2006; Pal, 2005; Rodriguez-Galiano *et al.*, 2012; Steele, 2000), and its application in rainfall retrievals is very new. Recently, Islam *et al.* (2014) used random forests to classify rainfall areas from satellite-borne passive microwave radiometers. At the same time, Kühnlein *et al.* (2014b) and Kühnlein *et al.* (2014a) investigated the potential of random forests as a tool within satellite-based rainfall retrievals using Meteosat Second Generation (MSG) Spinning Enhanced Visible and InfraRed Imager (SEVIRI) data. Both obtained promising results for the use of random forests in rainfall retrievals. Support vector machines are less frequently used in remote sensing (Mountrakis *et al.*, 2011) and have yet to be employed in optical rainfall retrievals. However, their potential has been shown in satellite-based land cover classifications (Kavzoglu and Colkesen, 2009; Pal, 2005) and in estimating biophysical parameters like chlorophyll concentration (Bruzzone and Melgani, 2005).

Though some rainfall retrieval techniques use different ML algorithms, to our knowledge, no study has compared different algorithms for rainfall assessment on the same dataset up until now. Hence, this study compares random forests (RF), neural networks (NNET), its extension averaged neural networks (AVNNET) and support vector machines (SVM) for their applicability in rainfall retrieval techniques.

This paper is structured as follows: Section 2.2 explains the methodology of the comparison study including data preprocessing, model training and the validation strategy. Section 2.3 presents the results of the comparison study which are then discussed in Section 2.4.

2.2 Data and methodology

Following the approach developed by Kühnlein *et al.* (2014b), rainfall area and rainfall rates were predicted for Germany during summer 2010. Day, twilight and night precipitation events were all treated separately due to differing information content about the cloud properties at different times of day. MSG SEVIRI data were used since they permit a quasi-continuous observation of the rainfall distribution and rainfall rate in near-real time. A radar-based precipitation product from the German Weather Service, RADOLAN RW (Bartels *et al.*, 2004), was used for ground truth data.

The general work flow included preprocessing the data to provide three datasets for model training: A day, twilight and a night dataset. The retrieval process was two-fold and consists of (i) the identification of precipitating cloud areas and (ii) the assignment of rainfall rates. Since the focus of this study is on the comparison of the algorithms, the validation of rainfall rate assignments was based on rainfall

areas derived from the radar network rather than the results from step (i). This ensures that the performance of rainfall rate models is comparable without confusion based on errors from the prior rainfall area delineation. Fig. 2.1 shows the work flow of the model training and comparison: For each dataset one model for rainfall area delineation and one model for rainfall rate assignment was tuned and trained for each of the chosen ML algorithms. The final models were applied to a test dataset and their performance was compared between the ML algorithms.

The following sections describe these steps in detail. All modeling and analysis was completed using the R environment for statistical computing (R Core Team, 2013). Model tuning, training and prediction was performed using the caret package (Kuhn, 2014a) as a wrapper package for a large list of machine learning algorithms implemented in R. Parallel processing was performed on 16 cores using the R package “doParallel” (Revolution Analytics and Weston, 2014).

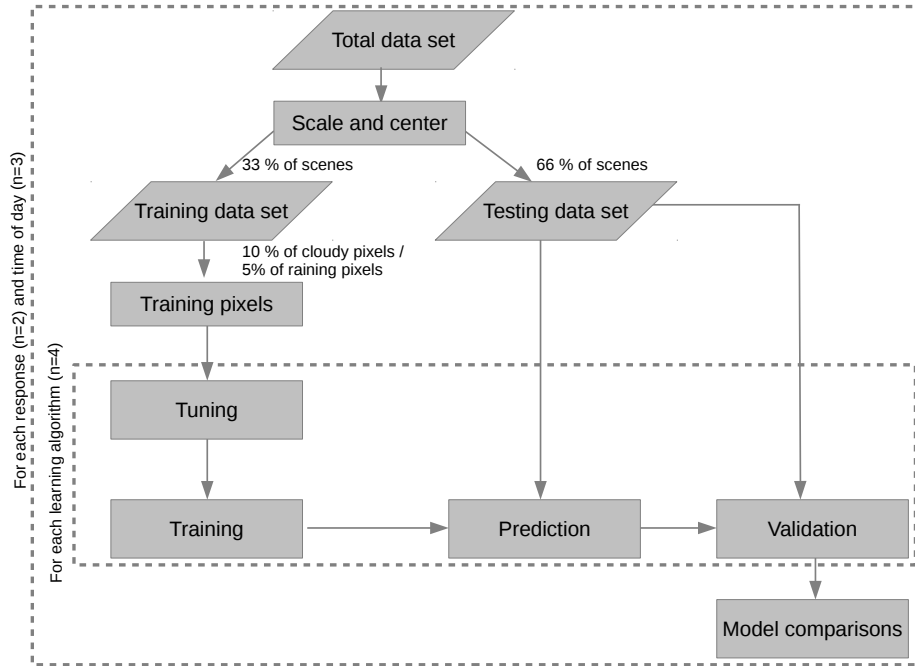


Figure 2.1: Flow chart of the main methodology applied in this study.

2.2.1 Datasets

2.2.1.1 Satellite data

MSG SEVIRI (Aminou *et al.*, 1997) scans the full disk every 15 minutes with a spatial resolution of 3 by 3 km at sub-satellite point. Reflected and emitted radiances are measured by 12 channels, three channels at visible and very near infrared wavelengths (between 0.6 and 1.6 μm), eight from near-infrared to thermal infrared wavelengths (between 3.9 and 14 μm), and one high-resolution

visible channel.

MSG SEVIRI data were downloaded from the EUMETSAT data center (www.eumetsat.int) and were preprocessed based on a newly designed Meteosat processing scheme implemented in co-operation with the computer science department at Marburg University. The processing chain uses `xxl` technology and custom raster extensions which were designed to support OpenCL acceleration (see <https://github.com/marburgedu/xxl>).

2.2.1.2 RADOLAN RW data

RADOLAN RW is based on measurements with a C-band Doppler radar of 16 German and neighboring radar stations. Rain intensity adapted Z-R relationships, statistical clutter filtering and shadowing effects are treated within an on-line calibration process. Furthermore, precipitation intensities are adapted with ground-based precipitation measurements. The precipitation product is available at a temporal resolution of one hour covering the entire area of Germany at a spatial resolution of 1 by 1 km (Bartels *et al.*, 2004).

2.2.2 Preprocessing of SEVIRI and Radar data

SEVIRI and RADOLAN RW data were preprocessed according to Kühnlein *et al.* (2014b) and afterwards available on an hourly basis. Scenes with at least 2000 rainy pixels were designated as precipitating events based on the RADOLAN RW product. The SEVIRI channels at visible and very near infrared wavelengths (0.6 to 1.6 μm) are not available at night. Use of the 3.9 μm channel is complicated during day and at twilight due to the varying solar component in this channel. Therefore, the dataset was split into day, twilight and night datasets. To ensure sufficient solar illumination in the VIS and NIR channels, scenes with a solar zenith angle less than 70° belong to the daytime dataset as suggested by Kühnlein *et al.* (2014a). Scenes with a solar zenith angle greater than 70° and less than 108° are assigned to the twilight, and those greater than 108° are assigned to the night dataset. The resulting daytime dataset consists of 327 scenes, the twilight dataset has 339 scenes and the night-time dataset has 130. The differences in the number of scenes for each period arises from a higher number of rain events during daytime and twilight conditions and from uneven ranges of solar zenith values used to separate into day, twilight and night. The distribution of rainfall areas was skewed with considerably more rain pixels than non-rain pixels (Tab. 2.1) in all three datasets. The overall mean rainfall rate was between 1.69 mm in the daytime dataset and 2.14 mm in the night dataset (Tab. 2.1). However, the mean values changed considerably between the scenes.

Table 2.1: *Summary of rainfall areas and rainfall rates of the three input datasets.*

	Day	Twilight	Night
	Rainfall area (% of cloudy pixels)		
Mean rain pixels	17.0	25.0	23.0
Mean non rain pixels	83.0	75.0	77.0
Min rain pixels per scene	2.8	2.4	1.5
Max rain pixels per scene	78.0	59.0	53.0
	Rainfall rate (mm)		
Mean rain rate	1.69	2.14	2.00
Standard deviation rain rate	2.15	2.51	2.12
Min mean rainfall rate per scene	0.56	0.67	0.44
Max mean rainfall rate per scene	7.35	6.91	4.38
Overall min	0.10	0.10	0.10
Overall max	168.00	72.20	40.66

2.2.3 Predictor variables

Summarizing the conceptual models of optical rainfall retrieval over the last several decades, optical cloud properties which are related to rainfall areas and rainfall rates are the cloud top height (CTH), the cloud top temperature (CTT), the cloud phase (CP) and the cloud water path (CWP). CTT-based retrievals have commonly used the CTT as a proxy for CTH based on the assumption that cold clouds produce (more) rain (e.g. Arkin and Meisner, 1987; Adler and Negri, 1988), which worked well for deep convective clouds but not for advective-stratiform systems. On the other hand, CWP-based retrievals take neither CTH nor CTT into account; they assume that precipitating clouds must have a large enough combination of the cloud droplets and the vertical extend of the clouds (e.g. Lensky and Rosenfeld, 1997; Nauss and Kokhanovsky, 2006; Thies *et al.*, 2008b,b; Kühnlein *et al.*, 2010). In accordance with previous studies as summarized by Kühnlein *et al.* (2014a), the spectral SEVIRI bands and derivations which can be used as proxies for these cloud properties were used as predictor variables (Tab. 2.2). The predictor variables contain the SEVIRI channels as well as channel combinations. Although this partially duplicates information, the channel combinations might be able to highlight patterns that are not apparent when only the individual channels are used. While Kühnlein *et al.* (2014a) also used all of these parameters, previous studies have generally been restricted to the utilization of a small subset and have generally focused on one of the two main conceptual models introduced above because of the parametric models used for the respective rainfall retrievals.

Table 2.2: *Overview of the predictor variables used to model rainfall areas and rainfall rates at day, twilight and night conditions (after Kühnlein et al., 2014b)*

Day	Twilight	Night
VIS0.6		
VIS0.8		
NIR1.6		
IR3.9	IR3.9	IR3.9
WV6.2	WV6.2	WV6.2
WV7.3	WV7.3	WV7.3
IR8.7	IR8.7	IR8.7
IR9.7	IR9.7	IR9.7
IR10.8	IR10.8	IR10.8
IR12.0	IR12.0	IR12.0
IR13.4	IR13.4	IR13.4
$\Delta T_{6.2-10.8}$	$\Delta T_{6.2-10.8}$	$\Delta T_{6.2-10.8}$
$\Delta T_{7.3-12.1}$	$\Delta T_{7.3-12.1}$	$\Delta T_{7.3-12.1}$
$\Delta T_{8.7-10.8}$	$\Delta T_{8.7-10.8}$	$\Delta T_{8.7-10.8}$
$\Delta T_{10.8-12.1}$	$\Delta T_{10.8-12.1}$	$\Delta T_{10.8-12.1}$
$\Delta T_{3.9-7.3}$	$\Delta T_{3.9-7.3}$	$\Delta T_{3.9-7.3}$
$\Delta T_{3.9-10.8}$	$\Delta T_{3.9-10.8}$	$\Delta T_{3.9-10.8}$
Sun zenith	Sun zenith	

2.2.4 Compilation of training and testing datasets

To avoid differences in algorithm performance resulting from different capabilities to deal with unscaled data, all predictor variables were centered and scaled. First, the data were visually checked for normal distribution. Scaling was then performed by dividing the values of the mean-centered variables by their standard deviations. The three datasets (day, twilight, night) were randomly split into a training and a testing dataset on a scene-basis. One-third of the scenes were used for training and two-thirds for validation. Models were trained on a pixel basis. Since the full set of training pixels would exceed acceptable computation times, only 5% of the cloudy pixels from the training scenes were selected for rainfall area training. For the training of rainfall rates, 10% of the rain pixels based on the RADOLAN RW product were considered. These were randomly chosen by stratified sampling to account for the distribution of the dataset.

2.2.5 Model tuning

For a description of the ML algorithms used, see James *et al.* (2013); Kuhn and Johnson (2013a). For all ML algorithms, a stratified 10-fold cross-validation was performed for a number of tuning values to determine the optimal model settings. To do this, the training samples were randomly partitioned into 10 equally sized folds with respect to the distribution of the response variable (i.e. raining cloud pixels, rainfall rate). Thus, every fold is a subset (1/10) of the training samples and has the same distribution of the response variable as the total set of training samples. Models were then fitted by repeatedly leaving out one of the folds. A model's performance was determined by predicting on the fold left out. The performance metrics from the hold-out iterations were averaged to the overall model performance for the respective set of tuning values. For classification models, the ROC-based distance to the perfect model (see section 2.2.5.4) (Fig. 2.2) was chosen as the performance metric. For regression models, RMSE was used.

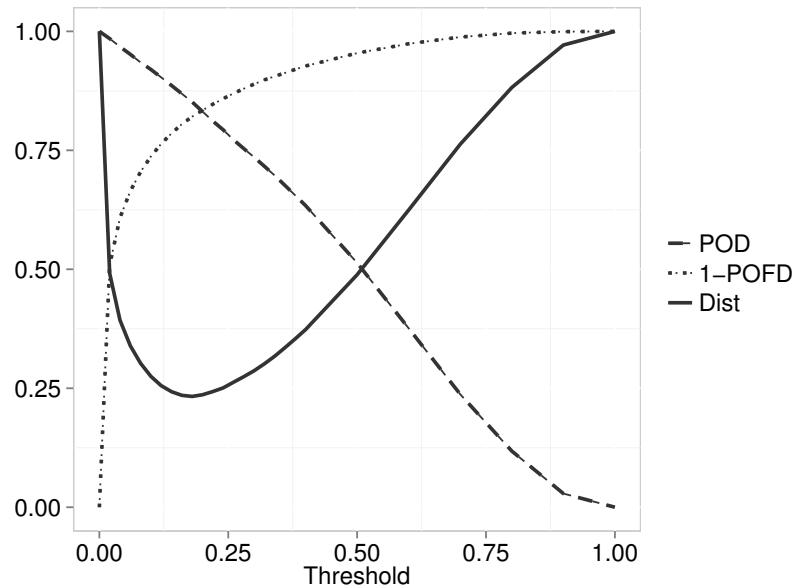


Figure 2.2: Example of threshold tuning. POD and POFD are calculated for each tuned threshold in the ROC space. The optimal operating point in ROC space leads to a classification with the best trade-off between failing to detect rain against the costs of raising false alarms. At this point, the distance to the perfect model (POD of 1 and a 1-POFD of 1) is smallest. In this case, classifying pixels with a probability for rain larger than 0.18 leads to best results.

2.2.5.1 Random forests tuning

The random forests implementation of the "randomForest" package (Liaw and Wiener, 2002) in R was applied. The number of predictor variables randomly

selected at each split (mtry) was tuned for each value between two and the number of input variables, following the suggestion of Kuhn and Johnson (2013a). The number of trees (ntree) was set to 1000 after no increase of accuracy was observed after 1000 trees.

2.2.5.2 NNET and AVNNET tuning

The NNET algorithm stems from the "nnet" package (Venables and Ripley, 2002) in R. The number of hidden units was tuned for each value between two and the number of predictor variables (Kuhn and Johnson, 2013a). Weight decay was tuned between 0 and 0.1 with 0.02 increments. The caret package offers support for averaging single neural nets. The tuning parameters used for NNET were kept the same for AVNNET. To average the model, five repetitions were performed with different random seeds.

2.2.5.3 SVM tuning

The "kernlab" package (Karatzoglou *et al.*, 2004) in R provided the SVM algorithm used in this study. The cost value was tuned for 0.5, 2, 8, 16, 32, 64, 128 and 512. A radial kernel function was used to account for non-linearity. Sigma was analytically solved as the median of $|x - x'|^2$ (Caputo *et al.* (2002) Karatzoglou *et al.*, 2004, in). For regression models, an ϵ -insensitive loss function (Vapnik, 1995) was used where $\epsilon = 1$ which controls the width of the tolerated error of data points around the hyper plane.

2.2.5.4 Threshold tuning for rainfall area classification models

The performance of ML algorithms suffers when training classes are highly unbalanced. In such a situation, ML algorithms tend to maximize performance by over-predicting the majority class (Liu *et al.*, 2006). This is particularly undesirable for this study's predictions as the intended prediction target (rain) usually represents the minority class. To overcome this problem, the optimal probability cut-off from predictive models was determined based on ROC analysis (Fawcett, 2006; Hamel, 2009) following the methodology of Kuhn (2014b). ROC curves as a metric for model selection describe a model's performance independently of the probability threshold which separates raining pixels from non-raining pixels. However, when probabilities are then translated to classes, the question of the most suitable threshold arises. Per default, a data point is classified as "rain" when its probability for rain is 0.5 or more. However, this default does not necessarily generate the best results. The optimal operating point in ROC space describes the threshold which leads to a classification with the best trade-off between sensitivity (POD) and specificity (1-POFD), i.e. between failing to detect

rain against the costs of raising false alarms. A perfect model has a POD of 1 and a POFD of 0. Calculating POD and POFD for different thresholds reveals the threshold where the distance to the perfect model is minimal (Fig. 2.2). For this reason, the threshold was treated as a tuning parameter for the selection of the most suitable model to predict rainfall areas. The optimal threshold was expected to be considerably smaller than 0.5 since rainy pixels were the minority class. Therefore, the range of thresholds tested went from 0 to 0.4 with increments of 0.02 and from 0.4 to 1 with increments of 0.1. The threshold with the minimal distance to a perfect model was used for the final training of the respective model.

2.2.6 Model prediction and validation

The tuning parameters (Tab. 2.3) that performed best were then applied to train the models. The trained models were used to predict rainfall areas and rainfall rates on the testing scenes. For validation, the performance measures described in the following sections were calculated on a scene-by-scene basis for each model.

Table 2.3: *Optimal tuning parameters which were the result of the tuning study and used for the final models.*

		Rainfall area			Rainfall rate		
		DAY	TWILIGHT	NIGHT	DAY	TWILIGHT	NIGHT
RF	mtry	10	9	11	5	14	7
	threshold	0.2	0.3	0.28			
NNET	decay	0.09	0.01	0.05	0.09	0.03	0.07
	size	18	14	14	14	8	14
	threshold	0.18	0.28	0.28			
AVNNET	decay	0.07	0.09	0.01	0.03	0.07	0.01
	size	12	12	14	18	14	14
	threshold	0.18	0.28	0.3			
SVM	sigma	0.0341	0.0525	0.0603	0.0352	0.0488	0.0589
	cost	32	512	128	512	128	32
	threshold	0.1	0.22	0.22			

2.2.6.1 Validation of rainfall area classification models

Categorical verification scores were calculated from confusion matrices showing agreement and disagreement between predicted and observed rainfall areas (Tab. 2.4, 2.5). Bias quantifies the over- or underestimation of raining pixels. Rain is underestimated in the model if the bias is <1 ; values >1 indicate overestimation of rain in the model. This is a pure quantitative measure that doesn't account for an agreement between observed and predicted rainfall areas. The

probability of detection (POD) gives the percentage of rain pixels that the model correctly identified as rain. The probability of false detection (POFD) gives the proportion of non-rain pixels that the model incorrectly classified as rain. Similar to POFD, the false alarm ratio (FAR) gives the proportion of predicted rain where no rain is observed. The critical success index (CSI) is the proportion of true positives to both kinds of errors. Correctly classified non rain pixels are not considered. This index is similar to the equitable threat score (ETS) which corrects for proper classification of rain just by chance. Hansen-Kuipers discriminant (HKD) and Heidke skill score (HSS) also account for chance agreement. HSS gives the proportion of correct classifications (both rain pixels and non-rain pixels) after eliminating expected chance agreement. HKD provides insight into whether predicting a pixel as rain leads to a considerable increase in false alarms. HSS is independent of bias in the classifications and is the difference between POD and POFD.

Table 2.4: *Confusion matrix as baseline for the calculation of the verification scores used for the validation of rainfall area predictions.*

		Observation	
		rain	no rain
Prediction	rain	True positives (TP)	False positives (FP)
	no rain	False negatives (FN)	True negatives (TN)

Table 2.5: *Calculation of the confusion matrix-based verification scores for the validation of rainfall area predictions.*

Name	Equation	Range	Optimum
Bias	$Bias = \frac{TP+FP}{TP+FN}$	$-\infty$ to ∞	1
Probability of Detection	$POD = \frac{TP}{TP+FN}$	0 to 1	1
Probability of False Detection	$POFD = \frac{FP}{FP+TN}$	0 to 1	0
False Alarm Ratio	$FAR = \frac{FP}{TP+FP}$	0 to 1	0
Critical Success Index	$CSI = \frac{TP}{TP+FP+FN}$	0 to 1	1
Equitable Threat Score	$ETS = \frac{TP-ph}{TP+FP+FN-ph}$ with $ph = \frac{(TP+FN)*(TP+FP)}{TP+TN+FP+FN}$	-1/3 to 1	1
Heidke Skill Score	$HSS = \frac{TP*TN-FP*FN}{[(TP+FN)*(FN+TN)+(TP+FP)*(FP+TN)]/2}$	$-\infty$ to 1	1
Hansen-Kuipers Discriminant	$HKD = \frac{TP}{TP+FN} - \frac{FP}{FP+TN}$	-1 to 1	1

2.2.6.2 Validation of rainfall rates regression models

Since most rainfall retrievals estimate on a 3-hour or 24-hour basis, the predictions and RADOLAN RW observations were aggregated and validated on a 3-hour and 24-hour basis. Because this study only includes scenes with at least 2000 rainy pixels and some of these scenes went into training, the aggregations (i.e. integral of the mean rainfall over 3 hours) were calculated from all available scenes that fell under the time interval (not necessarily 3 or 24 scenes). Error metrics as well as the coefficient of determination and relative error were used

for validation of the rainfall rate models. The mean error (ME) gives the mean of the signed differences between observed and predicted and is therefore useful for quantifying a bias and direction of the prediction. The mean absolute error (MAE) gives the strength of the prediction error. Root mean square error (RMSE) is more sensitive to large errors because it squares the errors and therefore penalizes large deviations from the mean. The difference between MAE and RMSE error gives valuable information about the variance of the errors. RMSE, ME and MAE are given in the same unit as the input, thus in mm. The coefficient of determination (R^2) gives the strength of the relation between measured rainfall rates and observed rainfall rates. The relative error is the rainfall rate error relative to the observed rainfall rate in % (Formula 2.1).

$$\frac{\sum_{i=1}^N |100 * (1 - \frac{\text{predicted}_i}{\text{observed}_i})|}{N} \quad (2.1)$$

2.3 Results

2.3.1 Comparison of predicted rainfall areas

Fig. 2.3 shows the performance of the rainfall area predictions on a scene basis. The differences in the models' performance were relatively low. The prediction of rainfall areas during daytime conditions yielded better results than during twilight and night-time conditions. Additionally, the variability between the scenes was lower during the day. All algorithms overestimated rainfall areas with a mean bias of 1.4 to 2.0 and a FAR of > 0.5 . For predicting rainfall areas during night conditions, RF had the highest bias. SVM generally featured the lowest bias values. On average, about 81% (SVM) to 85% (NNET) of the radar-observed rain pixels were also identified by the respective models during day. For all times of day, NNET and AVNNET showed the highest POD but also a relatively high FAR. Regarding AUC, CSI, ETS, HKD and HSS which consider both, POD and FAR, AVNNET and NNET showed the highest values (i.e. best performance). SVM performed the worst in terms of these scores but had a noticeably lower POFD than the other algorithms. NNET and AVNNET generally showed little difference, with AVNNET performing slightly better at twilight and night.

An exemplary scene from 2010/05/06 14:50 UTC (Fig. 2.4) was used to spatially depict the predictions. The spatial patterns of the rainfall area predictions which are shown in Fig. 2.5 are in accordance with the patterns indicated by the verification scores: All models show a high POD but considerably overestimated rainfall areas. SVM returned fewer false positives (no rain in reference, rain in the prediction) than the other models but also more false negatives (rain in reference, no rain in prediction) at the same time. AVNNET and NNET featured

the same spatial patterns of rain/no rain predictions. RF featured the highest FAR in this scene.

2.3.2 Comparison of predicted rainfall rates

Fig. 2.6 shows the performance of the rainfall rate predictions on a scene basis. The 3-hour prediction of rainfall rates during daytime conditions showed the lowest errors as well as the lowest variability. The differences in performance of the different algorithms were again very small. SVM showed lower MAE values than the other algorithms, regardless of the time of day. However, the RMSE was high, on average 2.13 mm. ME indicated a trend to overestimate rainfall rates using RF (average ME at daytime: 0.22). Neither NNET nor AVNNET had observable issues with over- or underestimation (ME: <0.05). SVM tended to underestimate rainfall rates (ME = -0.47). RMSE didn't change considerably between the models.

With aggregation on to 24 hours, R^2 increased significantly for all models. Aggregated to 3 hours, average R^2 scores were 0.37 (RF), 0.37 (NNET), 0.38 (AVNNET) and 0.34 (SVM). Aggregated to 24 hours, the R^2 increased to 0.42 (RF), 0.43 (NNET), 0.44 (AVNNET) and 0.39 (SVM). However, the RMSE increased at the same time compared to the 3-hour predictions for day and night. The relative error for RF, NNET and AVNNET was comparable. On average, the predicted rainfall rates deviated 57% from the observed values. SVM had a noteworthy high relative error of 123%.

The prediction of the exemplary day from 2010/05/06 (Fig. 2.7) showed that all algorithms underestimated high rainfall rates but overestimated low rainfall rates. This pattern was slightly more noticeable in AVNNET than in NNET. On average, SVM predicted lower rainfall rates than the other algorithms.

2.4 Discussion

2.4.1 Prediction of rainfall areas

The prediction of rainfall areas generally showed a moderate performance when compared to the RADOLAN RW data. Independent of the ML algorithm, the models had a high POD but noticeably overestimated rainfall areas, which led to the comparably low performance skills. This is because models were tuned to reach a good balance between POD and PFD, meaning that allowing for false positives was necessary to obtain a suitable proportion between POD and PFD. The findings highlight the general challenge of the retrieval to discern rain from non rain clouds, which needs further research. The edges of rain clouds,

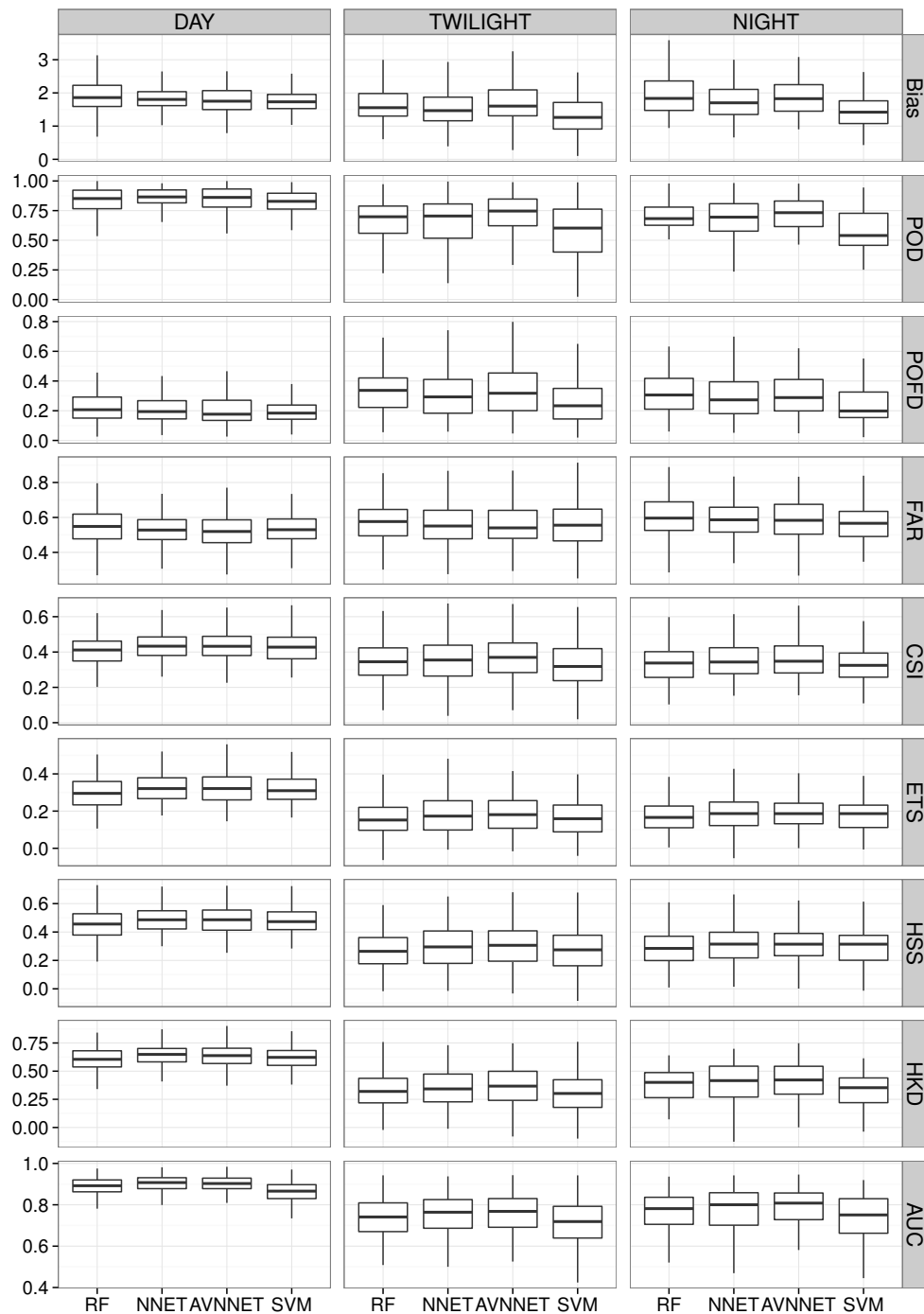


Figure 2.3: Comparison of the rainfall area prediction performances of the four ML algorithms at day, twilight and night conditions using the performance metrics Bias, probability of detection (POD), probability of false detection (POFD), false alarm ratio (FAR), critical success index (CSI), equitable threat score (ETS), Hansen-Kuipers discriminant (HKD), Heidke skill score (HSS) and area under the curve (AUC). Note that outliers are excluded since a visual assessment of the differences between models was impossible when a large span of values was illustrated.

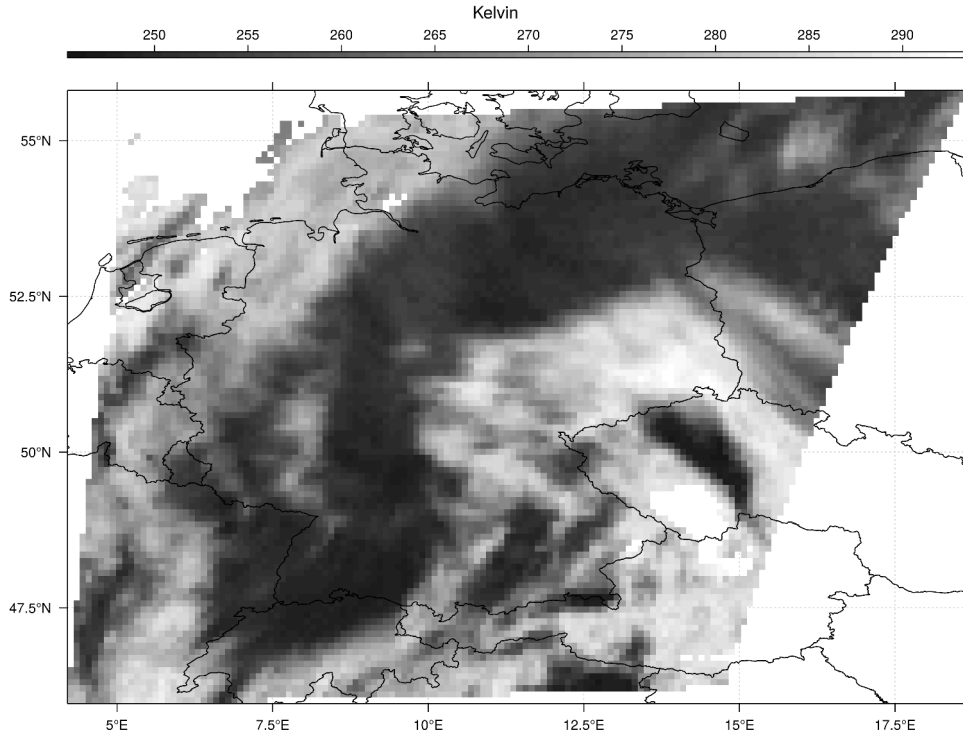


Figure 2.4: IR image of the MSG SEVIRI scene from 2010/05/06 14:50 UTC. This scene was used to visualize spatial patterns of rainfall predictions.

in particular, were often incorrectly predicted as raining. In this context, SVM was a very conservative classifier. Its predictions did lead to the lowest PFD but simultaneously yielded the lowest POD. No algorithm performed considerably better than the others. The threshold tuning appeared to be a valuable approach since it yielded better results than the static approach of Kühnlein *et al.* (2014a) to account for the unbalanced data.

2.4.2 Prediction of rainfall rates

AVNNET was the best algorithm for rainfall rate assignment. It performed slightly better than NNET especially in twilight and night conditions. Both datasets were smaller so that averaging of single NNET predictions might have been advantageous in cases where only limited data are available for training. The fact that AVNNET and NNET are similar in performance indicate the stability of single NNET models since averaging only slightly improves the performance. However, averaging single NNET predictions leads to a generalization which becomes obvious in less extreme values, neither low nor high rainfall rates. RF, similar to AVNNET, averages single predictions (from each tree) and therefore tend to overestimate low rainfall rates and underestimate high rainfall rates.

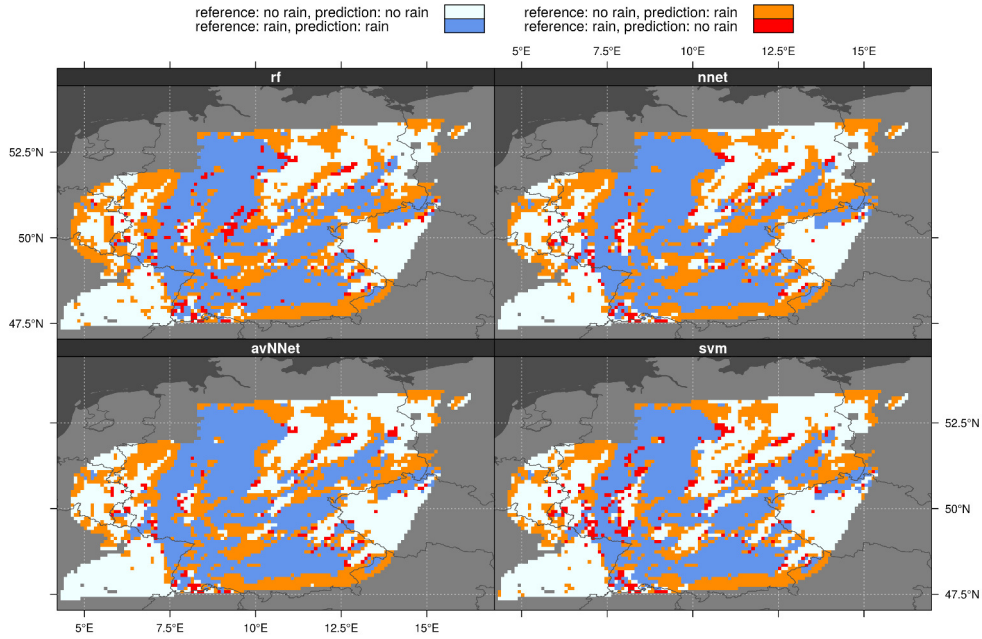


Figure 2.5: Visualization of the rainfall area predictions of the four ML algorithms for the exemplary scene from 2010/05/06 14:50 UTC. Green pixels were correctly predicted either as rain (dark green) or no rain (light green). Red and orange colors represent prediction errors, either due to false alarms (orange) or misses (red).

Therefore, each model is suited to different characteristics and should be chosen depending on what must be modeled. For modeling extreme values, for example, ensemble methods such as AVNNET and RF are not suitable.

Generally, predicting rainfall rates with daytime conditions showed better results than the twilight and night predictions which supports the findings of Kühnlein *et al.* (2014b). This might be because of the additional information about CWP which is directly related to solar scattering processes assessable by VIS and NIR channels that are not available at twilight or night. Each algorithm reflected the same behavior according to temporal aggregation and time of day.

A temporal aggregation of rainfall rates to 24 hours increased the performance of the retrieval technique. Note that since not all scenes of one day went into prediction and aggregation, performance could be expected to improve when a full set of one day was available. Aggregated on 24 hours, the performance of NNET and AVNNET converges. Differences in model performances were not noticeable between the 3-hour and 24-hour predictions.

An explicit comparison to other rainfall retrievals was neither in the scope of this study nor directly possible due to differences in ground truth data, spatial extent, time period as well as spatial and temporal resolution. However, readers who are not familiar with the range of rainfall retrieval performances can refer to

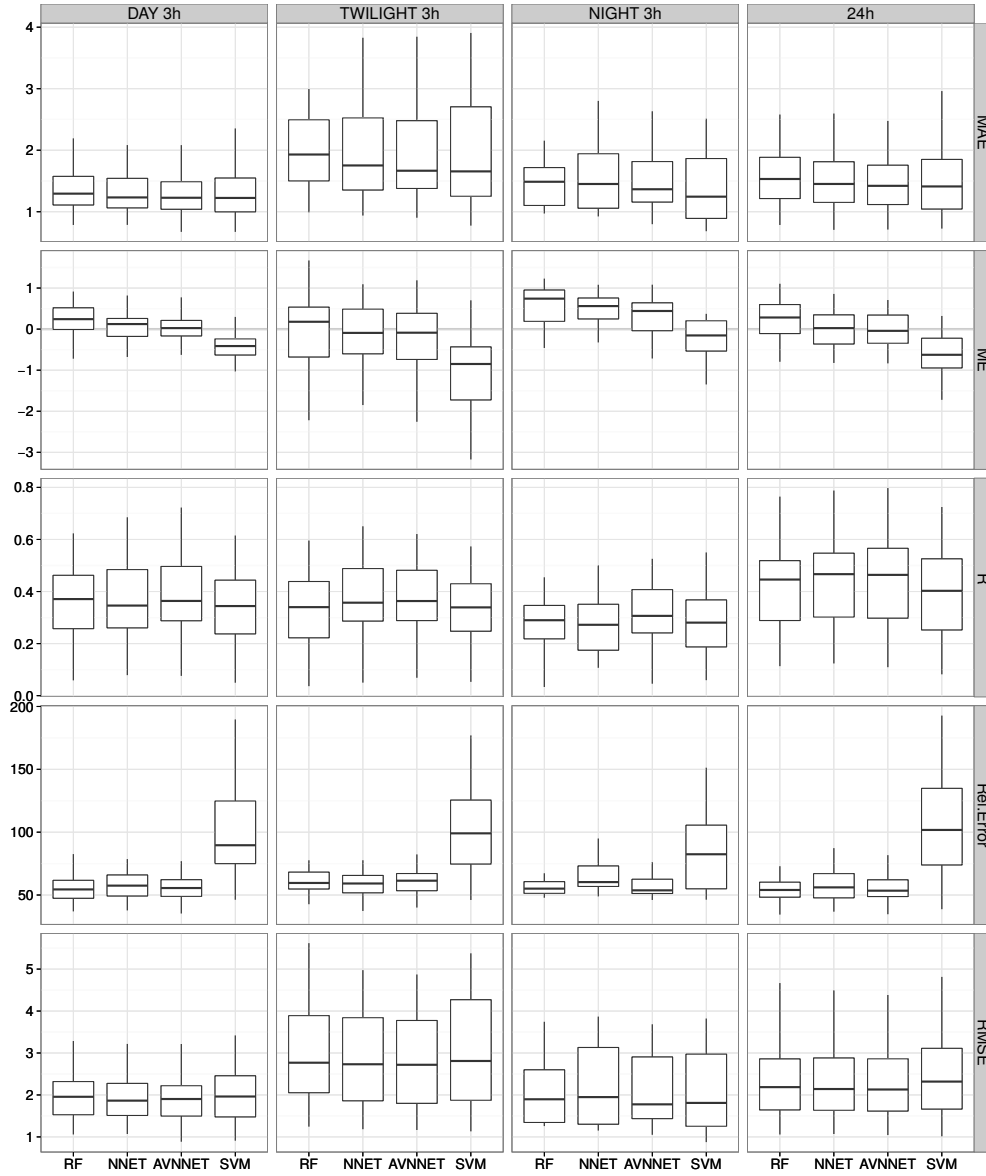


Figure 2.6: Comparison of the rainfall rate prediction performances of the four ML algorithms on a scene-by-scene basis aggregated to 3 hours as well as 24 hours. The performance metrics are mean absolute error (MAE), mean error (ME), coefficient of determination (R^2), relative error (Rel. Error) and root mean square error (RMSE). Note that outliers are excluded since a visual assessment of the differences between models was impossible when a large span of values was illustrated.

Kidd and Levizzani (2011) who compared different precipitation products over Northwest Europe with a spatial resolution of $25^\circ \times 25^\circ$. Compared to the performance of the retrievals included in their study, our retrieval can at least keep up with even complex retrievals like TRMM products or CMORPH even despite the higher spatial resolution.

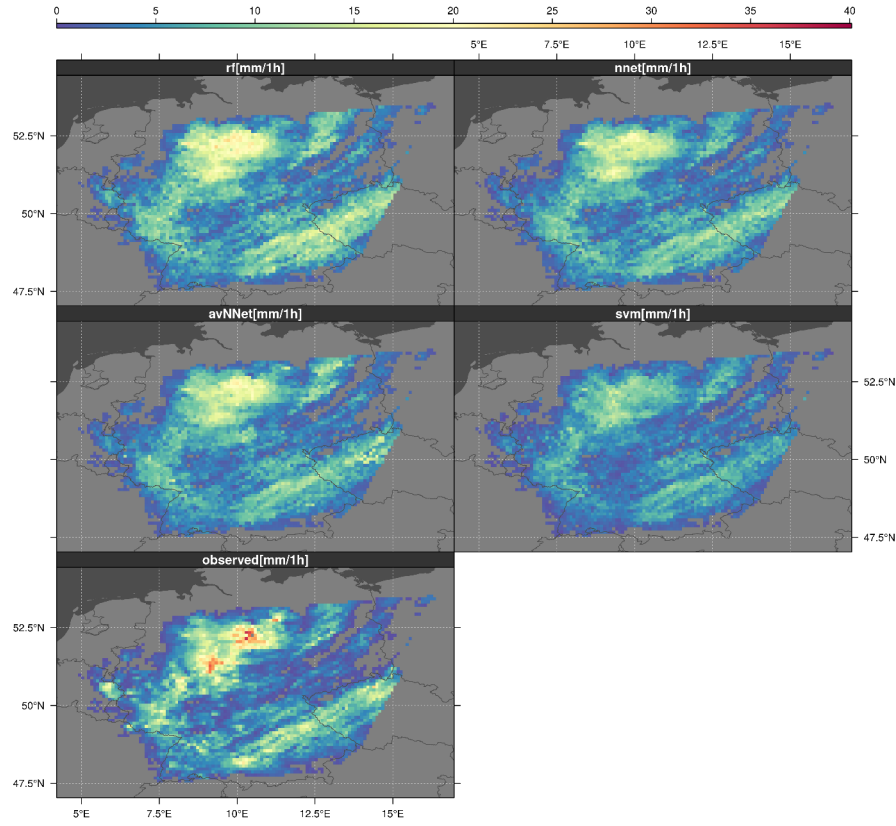


Figure 2.7: Visualization of the 24-hour aggregated rainfall rate predictions of the four ML algorithms for the exemplary day 2010/05/06.

2.4.3 Technical considerations

Though recent developments in parallel computing make it possible to train ML models based on large datasets, processing time is still a decisive factor when choosing algorithms, especially when dealing with datasets that have high spatial and temporal resolution, as in the case of optical rainfall retrievals. Tab. 2.6 depicts the computation times for this study using the example of the day dataset. Model training with optimal tuning parameters was fastest with NNET (three minutes) and slowest with SVM (108 minutes). In addition to being 30 times faster than SVM, NNET was also three times faster than RF in the classification models. Due to averaging multiple models, AVNNET was two to three times slower than NNET. However, the overall training contained an extended tuning study, which differed in length due to each of the four algorithm's different requirements. Since tuning is an essential step in model building, the time it takes to complete is also valuable information to take into consideration. In regression mode (i.e. rainfall rate prediction), tuning considerably extended the computation time for SVMs - to up to 43 hours. Though NNET uses one additional tuned

parameter, it was still hands down the fastest algorithm needing only 8 minutes for the same task.

Using caret as a wrapper package, all models essentially require the same number of lines of code. Efforts for customization of each model are essentially non-existent.

Table 2.6: *Processing time in minutes for model tuning and training as well as just model training with the optimal tuning parameters. The values are based on the models for daytime, which contained the largest number of training scenes. The number of pixels used for training rainfall areas was 88,751 (5% of cloudy pixels) and 28,251 (10% of rain pixels) for training rainfall rates.*

	Rainfall area models		Rainfall rate models	
	Tune and Train	Train with optimal	Tune and Train	Train with optimal
RF	80	11	358	34
NNET	46	3	9	1
AVNNET	135	9	41	3
SVM	718	108	2600	2076

2.5 Summary and conclusions

This study investigated the suitability of different ML algorithms for their applicability in optical rainfall retrievals. Though the algorithms showed very little difference in their performance, NNET proved to be the most suitable algorithm. On average it performed best in rainfall area delineation as well as rainfall rate assignment. Its comparably fast computation time is another advantage when working with large datasets that are commonplace in remote sensing based rainfall retrievals. In some cases AVNNET performed better than single NNET, however, the slight increase in performance is not justified by the considerably longer computation time. RF showed an intermediate performance with intermediate computation time. However, it is worth noting that among the algorithms tested, RF is the most intuitive and easy to use since it does not require data to be preprocessed and may therefore be justified despite its slightly lower performance. SVM performed worst and required far too much computation time. We therefore recommend using NNET as an ML algorithm in optical rainfall retrieval applications. Since (AV)NNET was the best performing algorithm in both, rainfall area detection and rainfall rate assignment, there is no need to use different algorithms for the two steps in the rainfall retrieval.

Despite our recommendation, no algorithm performed considerably better than the others. Thus, continued research is recommended determining suitable predictors for rainfall, rather than in optimizing through the choice of an ML algorithm. So far, our work has solely focused on the spectral properties of clouds, which is, according to Thies and Bendix (2011), most important for applications in cloud classification. However, we assume that textural properties are highly related to rainfall rates, as they present spatial characteristics of a cloud. This assumption is supported by a recent study of Giannakos and Feidas (2013) which shows the potential of textural parameters for the classification of rainy clouds. Thus, we will employ spatial filters across different SEVIRI channels as proxies for cloud texture and larger scale patterns at the same time in future studies on optical rainfall retrievals. Using these variables we will expand on the contemporary purely pixel-based approach by including information about their neighborhood.

Acknowledgments This work was financially supported by the Federal Ministry of Education and Research (BMBF) within the IDESSA project (grant no. 01LL1301) which is part of the SPACES-program (Science Partnership for the Assessment of Complex Earth System processes). The authors are grateful to the German Weather Service (DWD) for providing the RADOLAN RW and the EUMETSAT Earth Observation Portal (<https://eoportal.eumetsat.int/>) for providing the Meteosat dataset. We would further like to thank Max Kuhn for helping adapt his work on threshold tuning to the ML algorithms used in this study.

Chapter 3

Revealing the potential of spectral and textural predictor variables in a neural network-based rainfall retrieval technique

Hanna Meyer (1), Meike Kühnlein (1),
Christoph Reudenbach (1), Thomas Nauß (1)

(1) Faculty of Geography, Philipps-University Marburg, Deutschhausstr. 10, 35037 Marburg, Germany

Published in

Remote Sensing Letters, 2017, 8, 647–656

Received 29 Nov 2016

Accepted 16 Mar 2017

Published online: 05 Apr 2017

3 Revealing the potential of spectral and textural predictor variables in a neural network-based rainfall retrieval technique

Hanna Meyer, Meike Kühnlein, Christoph Reudenbach, Thomas Nauß

Abstract

Estimating rainfall areas and rates from geostationary satellite images has the opportunity of both, a high spatial and a high temporal resolution which cannot be achieved by other satellite-based systems until now. Most recent retrieval techniques are solely based on spectral channels of the satellites. These retrievals can be classified as "purely pixel-based" because no information about the neighbourhood pixels is included. Assuming that precipitation is highly correlated with cloud processes and therefore with cloud texture, textural information derived from the neighbourhood of a pixel might give valuable information about the cloud type and hence about a respective probability of the rainfall rate. To study the potential of textural variables to improve optical rainfall retrieval techniques, rainfall areas and rainfall rates were estimated over Germany for the year 2010 using a neural network approach. In addition to the spectral predictor variables from Meteosat Second Generation (MSG), different Grey Level Co-occurrence Matrix (GLCM) based textural variables were calculated from all MSG channels. Using recursive feature selection, models were trained and their performance was compared to spectral-only models. Contrary to the expectations, the performance of the models did not increase when textural information was included.

3.1 Introduction

Estimating rainfall from geostationary satellite images has the opportunity of both, a high spatial and a high temporal resolution which cannot be achieved by other satellite-based systems until now. Though recent studies indicate the great potential of optical rainfall retrievals (see valuable overviews by Kidd and Levizzani, 2011; Levizzani *et al.*, 2002, 2001; Prigent, 2010; Thies and Bendix, 2011), the task to accurately estimate rainfall from space remains challenging due to the variability of rainfall patterns on the one hand, and due to its high spatio-temporal dynamic on the other.

The majority of recent optical rainfall retrievals are based on machine learning algorithms to relate the spectral satellite information to rainfall areas or rainfall rates rather than using parametric approaches (Capacci and Conway, 2005; Grimes *et al.*, 2003; Rivolta *et al.*, 2006; Giannakos and Feidas, 2013; Hsu *et al.*, 1997; Kühnlein *et al.*, 2014a). Though different machine learning algorithms are used in the respective retrieval techniques, Meyer *et al.* (2016b) compared the performances of different ML algorithms and concluded that there is a need to improve retrievals by defining suitable predictor variables rather than optimizing the retrievals by the choice of the ML algorithm which all performed similar if trained properly.

Among the recent retrievals, the predominating predictor variables that are used are infrared channels (Feidas and Giannakos, 2012; Behrangi *et al.*, 2009b) which are in some studies complemented by visible, near infrared and water vapour channels as well as various channel differences (Kühnlein *et al.*, 2014a,b; Thies *et al.*, 2008b,c; Ba and Gruber, 2001). These retrievals can be classified as "purely pixel-based" because each pixel in an image is treated completely independent from its neighborhood and no information about the surrounded pixels is included. However, textural information derived from the neighbourhood of a pixel might give valuable information about the cloud type and, due to corresponding microphysical processes, about a respective probability of the rainfall rate. Related to this, textural measures of cloud surfaces were repeatedly used as proxy for the cloud type (Christodoulou *et al.*, 2003; Ameur *et al.*, 2004; Welch *et al.*, 1988; Giannakos and Feidas, 2013), see also a review by Tapakis and Charalambides (2013). For example, Kidd and Levizzani (2011) describe stratus clouds as appearing smooth in a certain visible environment while convective clouds tend to have a heterogeneous surface in the visible as well as in the infrared (Christodoulou *et al.*, 2003).

Grey Level Co-occurrence Matrix (GLCM) based textural measures by Haralick *et al.* (1973) indicate the spatial distribution of grey values in a specific environment and are commonly used in remote sensing of clouds. Welch *et al.*

(1988) and Christodoulou *et al.* (2003) used the GLCM based textural metrics for cloud classification using Landsat and Meteosat Second Generation (MSG) Spinning Enhanced Visible and Infrared Imager (SEVIRI) images respectively. Some studies also successfully included GLCM based textural parameters in optical rainfall retrieval techniques (Uddstrom and Gray, 1996; Liu *et al.*, 2014; Giannakos and Feidas, 2011, 2012; Hong *et al.*, 2004; Hsu *et al.*, 1997). However, in these retrievals, texture is solely derived from IR channel brightness temperatures in a 3×3 pixel window, or/and the study are based on a very limited number of scenes. An extended study on the contribution of textural information in different pixel environments using the full spectral information of optical satellite data is still lacking.

Given the still challenging character of satellite-based rainfall estimation, the aim of this study is to analyze the potential of textural variables in different spectral ranges for an improvement of optical rainfall retrieval techniques.

3.2 Methods

Rainfall areas and rainfall rates were estimated over Germany for the year 2010. Therefore, MSG SEVIRI data were used since they permit a quasi-continuous observation of the rainfall distribution and rainfall rate in near-real time. A radar-based precipitation product from the German Weather Service, RADOLAN RW (Bartels *et al.*, 2004), was used for ground truth data. The general retrieval process was two-fold and consists of (i) the identification of precipitating cloud areas and (ii) the assignment of rainfall rates. Since the focus of this study is on revealing the potential of textural variables, the assignment of rainfall rate was based on rainfall areas derived from RADOLAN RW rather than from the results from step (i). This ensures that the performance of rainfall rate models is comparable without confusion based on errors resulting from the rainfall areas delineation step. Due to unavailability of visible channels during night, rainfall areas and rates were modeled for day and night scenes separately. For all modeling tasks (rainfall areas during day, rainfall areas during night, rainfall rate during day, rainfall rate during night), the individual models were compared based on the utilization of spectral and textural variables or spectral variables only (Fig. 3.1). The following sections describe the model comparison workflow in detail. Modeling and analysis was completed using the R environment for statistical computing (R Core Team, 2015) in conjunction with the caret package (Kuhn, 2014a) for machine learning applications.

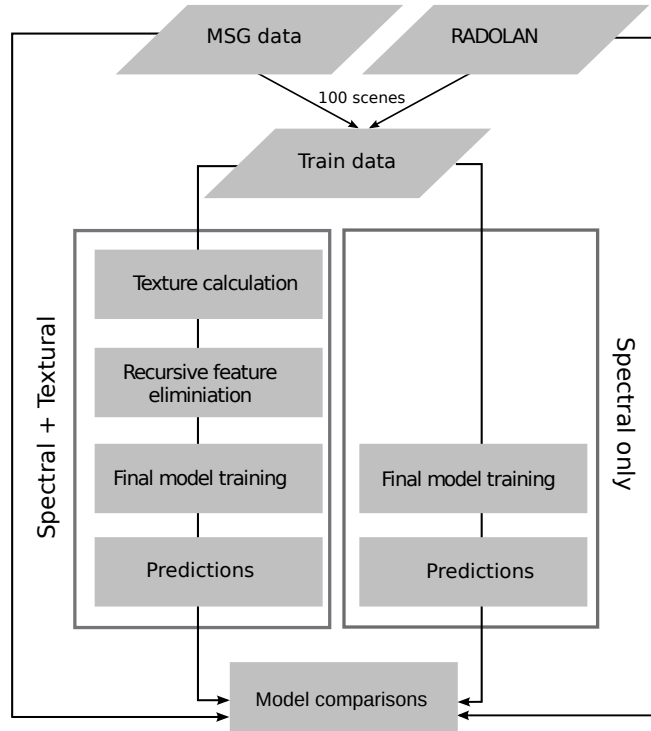


Figure 3.1: Overview of the methods to compare models that use spectral and textural variables with models that use spectral variables only.

3.2.1 Satellite and ground truth data

MSG SEVIRI (Aminou *et al.*, 1997) scans the full disk every 15 minutes with a spatial resolution of 3×3 km at sub-satellite point. The spatial resolution in the study area Germany is approximately 4.5×4.5 km. Reflected and emitted radiances are measured by 12 channels, three channels at visible and very near infrared wavelengths (between 0.6 and $1.6 \mu\text{m}$), eight channels ranging from near-infrared to thermal infrared wavelengths (between 3.9 and $14 \mu\text{m}$) and one high-resolution visible channel. In this study, SEVIRI data from the year 2010 was preprocessed and a cloud-mask based on Kühnlein *et al.* (2014b) was computed on an hourly basis. All MSG SEVIRI channels except for the high-resolution visible channel were included in this study for modelling rainfall during day. During night, the three channels in the visible and near infrared were not used since they don't provide reliable information. In addition to the spectral channels, combinations of brightness temperature differences, as for example the difference between the $6.2 \mu\text{m}$ and the $10.8 \mu\text{m}$ channel ($T_{6.2} - T_{10.8}$), were calculated following Kühnlein *et al.* (2014a) and resulting in 17 spectral variables during day (spectral channels and the channel differences $T_{6.2} - T_{10.8}$, $T_{7.3} - T_{12.1}$, $T_{8.7} - T_{10.8}$, $T_{10.8} - T_{12.1}$, $T_{3.9} - T_{7.3}$, $T_{3.9} - T_{10.8}$) and 14 spectral variables during night. The GLCM based metrics homogeneity, contrast, dissimilarity, entropy and second moment (Haralick *et al.*, 1973) were calculated from all spectral predictors

in a 3×3 as well as a 5×5 pixel environment using the "glcm" package in R (Zvoleff, 2015). To avoid high computational efforts, the number of quantization/grey levels was reduced from 1024 to 128. The GLCM was calculated for four directions and the averages of all directions were taken as final parameters. In addition to the GLCM based indices, the mean, minimum, maximum and standard deviation values in the 3×3 as well as 5×5 pixel environment of all spectral variables were used as further predictors. In total, including the spectral variables, 342 potential predictor variables during day and 266 during night were provided for the modelling approach. All predictors were centered and scaled by dividing the values of the mean-centered variables by their standard deviations.

RADOLAN RW data was used as ground truth. It is based on measurements with a C-band Doppler radar of 16 radar stations covering the area of Germany. Relationships between radar reflectivity and precipitation rate (Z/R relationship), statistical clutter filtering and shadowing effects are treated within an on-line calibration process. Furthermore, precipitation intensities are adapted with ground-based precipitation measurements. The precipitation product is available at a temporal resolution of one hour covering the entire area of Germany at a spatial resolution of 1×1 km (Bartels *et al.*, 2004). RADDOLAN RW data were re-projected to the geostationary projection using a bi-linear resampling in order to match the geometry of the SEVIRI data.

3.2.2 Compilation of training and test data sets

Scenes with at least 3000 rainy pixels in the RADOLAN RW product were included in the study. Only these scenes were considered for further analysis. The scenes were split into day scenes (scenes with a solar zenith angle less than 70°) and scenes where the visible channels are not reliably available (i. e. night and twilight). All twilight and night scenes are treated equally in this study and are termed "night" in the following. Since model training using several hundred of predictor variables is computing cost intensive, a selection of 100 training scenes during day and night were each selected randomly. The random selection gave 100 rain events from 78 different days during day and from 76 different days during night. All other rain events in the year 2010 were used for independent testing. From the training sample, 5% of the cloudy pixels were considered for training the rainfall area models and another 25% were used for the rainfall rate training. The selection of training pixels was performed using stratified random sampling to account for the distribution of the dataset. The final training sample size consisted of 110920 pixels for rainfall areas training during day and 63896 during night as well as 141931 pixels for rainfall rate training during day and 102384 during night.

3.2.3 Neural network training

Neural networks are a well-established method in cloud and rainfall detection studies (e.g. Hsu *et al.*, 1997; Hong *et al.*, 2004; Lazri *et al.*, 2014; Tebbi and Haddad, 2016) and it was shown that they perform best in this optical rainfall retrieval technique with a high computation speed which is important considering the high amount of predictors and data points (Meyer *et al.*, 2016b). We used a single-hidden-layer neural network, implemented in the "nnet" package (Venables and Ripley, 2002) in R. Neural networks require two hyperparameters to be tuned: the number of neurons in the hidden layer and the weight decay (Kuhn and Johnson, 2013a). For all steps of parameter tuning as well as variable selection and model training, a stratified 10 fold cross-validation was performed to determine the optimal model settings. Thus, the training samples were randomly partitioned into 10 equally sized folds with respect to the distribution of the response variable (i. e. equal distribution of rainy/non rainy cloud pixels and equal distribution of rainfall rates respectively). Models were then fitted by repeatedly leaving one of the folds out. Performance of a model was determined by predicting on the respective held-out fold. Receiver Operating Characteristics (ROC) were used as performance metric for rainfall areas models and coefficient of determination (R^2) for rainfall rate models. The performance metrics from the hold-out iterations were averaged to the overall cross validated model performance for the respective set of tuning values.

The performance of ML algorithms suffers when training classes are highly unbalanced since they tend to maximize performance by over-predicting the majority class. This is particularly critical for the modeling task of this study as the intended response variable (rainy clouds) usually represents the minority class. To overcome problems caused by unbalanced classes (see e. g. Liu *et al.*, 2006), the optimal probability cut-off from rainfall area models was determined based on ROC analysis (Fawcett, 2006; Hamel, 2009) following the methodology of Kuhn (2014b). We therefore used the threshold from the estimated probabilities as additional tuning parameter in the classification models (tuned between 0.0 and 0.4 with increment 0.02 and between 0.5 and 1.0 with increment 0.1). The threshold leading to the minimal distance to a perfect model was used for the final training of the respective model (See Meyer *et al.* (2016b) for further description on this method).

3.2.3.1 Recursive feature selection

Though neural networks are known as being able to deal with highly correlated predictor variables, Kuhn and Johnson (2013a) have shown that neural networks are not as unaffected by adding non-informative or redundant parameters. Further, from a technical point of view, many predictors result in a high amount of

data which increases computation costs. Feature selection is a suitable method to take all potential predictor variables into account but overcome the mentioned issues. We used recursive feature elimination according to Guyon *et al.* (2002) which is implemented in the caret R package (Kuhn, 2014a). Recursive feature elimination fits the model first with all predictor variables. It then calculates variable importance according to the weights method of Gevrey *et al.* (2003) and removes the least important variables. In the next step, the model is re-calculated with the reduced number of variables. This step is repeated for different numbers of variables. The best number and combination of variables is then determined by comparing the performance of the individual models.

Since feature selection is very computation time consuming, hyperparameter tuning was reduced to a minimum: The number of neurons in the hidden layer was tuned between 2 and 10 with increment 2; 15 to 30 with increment 5; 40 to 80 with increment 10 and 100 to the number of predictor variables with increment 50. Weight decay was kept constant at 0.05. A more detailed tuning study was carried out after the optimal variables were determined.

3.2.3.2 Fine tuning and model training

Models using the optimal variables determined by feature selection as well as the spectral-only models were extensively tuned and trained. Weight decay was tuned between 0 and 0.1 with increment 0.02. The number of neurons in the hidden layer was tuned between 2 and the number of predictor variables with increment 2. The best performing tuning parameters were applied for final model training. The trained models were used to estimate rainfall areas and rainfall rates of the testing scenes.

3.3 Results

The performance of the rainfall areas models first increased with the number of predictor variables, for both, day and night (Fig. 3.2a). The optimal number of predictor variables was identified to be 20 ($\text{ROC} = 0.902$) for day and 14 for night ($\text{ROC} = 0.786$). Regarding the night model, the performance then dropped down and remained constant from 75 variables onward. The day model was not affected by a reduced performance with an increasing number of predictor variables. The performance remained constant after the optimal ROC value was reached. The rainfall rate models were more affected by the number of predictor variables (Fig. 3.2b). The performance first increased to its maximum using 30 variables during day ($R^2 = 0.313$) and 14 variables during night ($R^2 = 0.211$) respectively. The R^2 then rapidly decreased in both, the day and the night models.

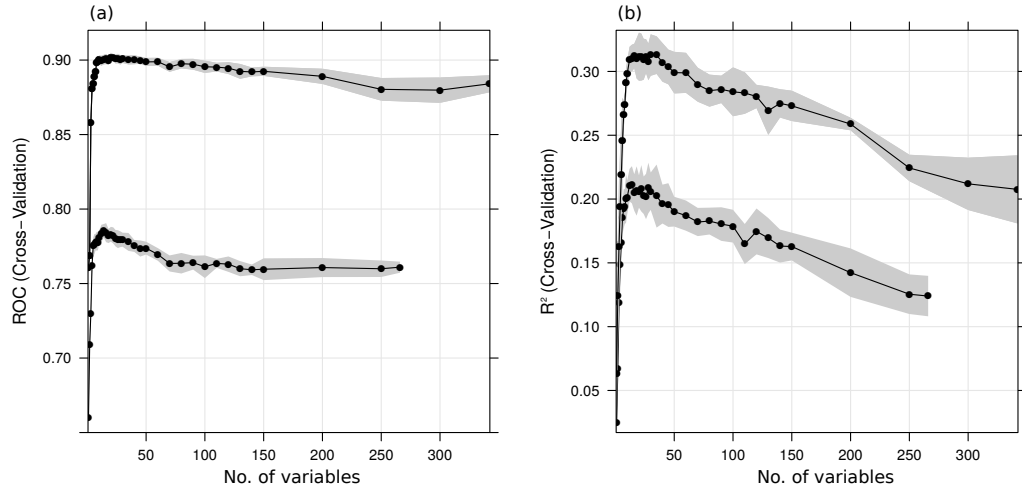


Figure 3.2: *Dependence of the number of variables on the performance of (a) rainfall areas during day (upper line) and night (lower line) indicated by ROC and (b) rainfall rates during day (upper line) and night (lower line) indicated by R^2 . The grey areas show the standard error.*

There were no significant differences between the models which used the optimal variables revealed during the (textural) feature selection and the models which used spectral variables only (Fig. 3.3). The average Root Mean Square Error (RMSE) of rainfall rate estimation during day was 1.09 mm for spectral-only models and 1.08 mm for the spectral+textural model. During night, the average RMSE was 1.02 in both models. Regarding the estimation of rainfall areas, both models had a Probability Of Detection (POD) of 0.70 and a Probability Of False Detection (POFD) of 0.36 during night. During day, the POFD of both, spectral-only as well as spectral+textural models was 0.20. POD was 0.80 in the spectral-only model and 0.81 in the spectral+textural model.

3.4 Discussion and Conclusion

The data-driven methodology allowed to initially include a wide range of potential predictors including even correlated and potentially uninformative variables. The decreasing performance with increasing number of variables, however, showed that a feature reduction is necessary when a high number of predictor variables is presented to the models. In general, the performance of the retrieval is in the same range as indicated by similar studies (Kühnlein *et al.*, 2014b,a; Giannakos and Feidas, 2013). Surprisingly, the use of textural variables did not considerably increase the performance of the models. Concerning the delineation of rainfall areas, these findings correspond to those of Giannakos and Feidas (2011). Though Giannakos and Feidas (2012) showed that textural variables can slightly improve estimations of rainfall rate delineations compared to spectral-only mod-

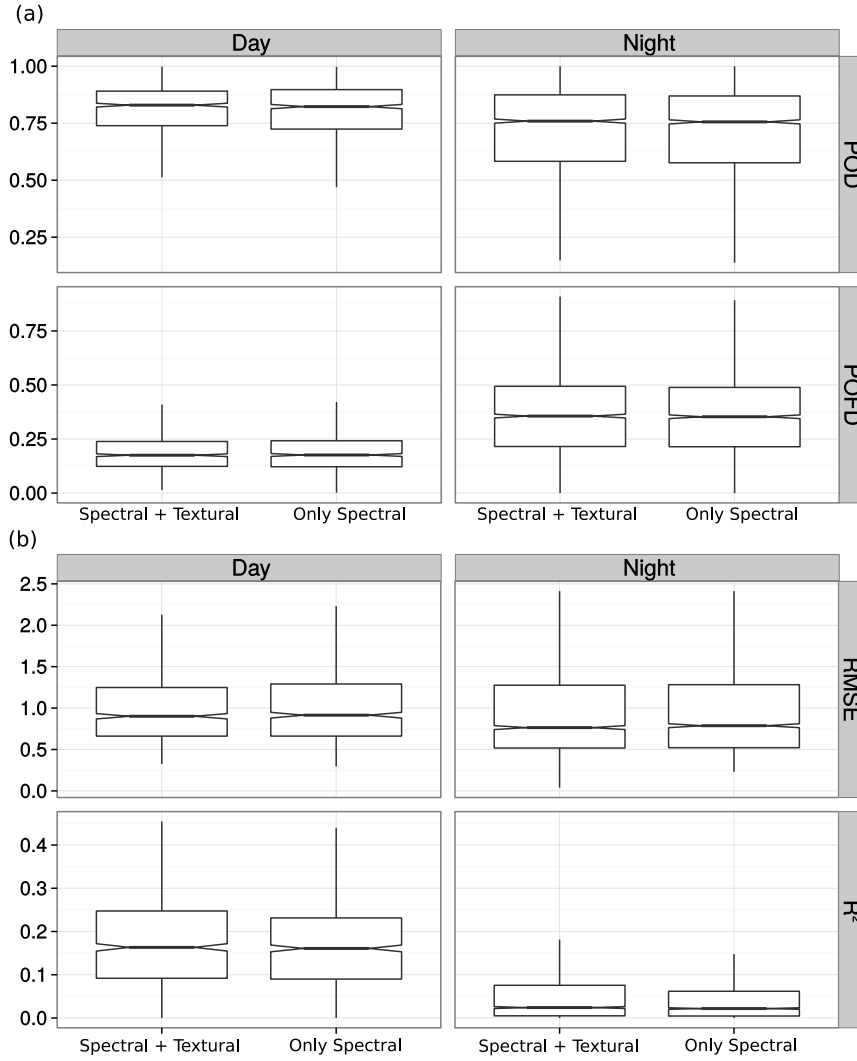


Figure 3.3: Boxplots showing the performance of the full models as well as spectral-only models (a) for rainfall areas indicated by POD and POFD and (b) for rainfall rates indicated by RMSE and R^2 . Each data point in the boxplot corresponds to one MSG scene of the test data set. Note that outliers of the RMSE are excluded to facilitate a visual assessment of the differences between models.

els, these findings could not be confirmed by this study. The contradictions might result from the considerable smaller number of training scenes used by Giannakos and Feidas (2012) which induce a higher risk for overfitting.

Though during feature selection many textural variables were selected for the final model, the independent model validation did not indicate an improvement compared to the spectral-only models. The differences between the cross validated performance and the performance indicated by independent model validation might be a matter of slight overfitting. Since the cross validation was not based on a leave-one-scene-out cross validation, the samples have not been independent which might explain the difference between cross validated performance

and the completely independent performance.

No advantage of adding textural variables to spectral predictor variables could be found by using the presented simple neural network in conjunction with recursive feature selection. However, it can't be excluded that other feature selection strategies like e.g. principal component analysis (see Behrangi *et al.*, 2009a, for its application in rainfall retrievals), auto-associative neural networks (Kerschen and Golinval, 2004) or forward feature selection in conjunction with the neural network (Meyer *et al.*, 2016a) would be able to make sense of the textural information. Also, though using single-layer neural networks is common practice for rainfall retrievals (e.g. Giannakos and Feidas, 2013), the amount of potential predictor variables in this study is exceptionally high. Against the background of extremely high correlations between the predictor variables it can't be excluded that deeper architectures of neural networks would be able to detect relations that could not be revealed using a single hidden layer. Therefore, increasing the number of hidden layers (Grimes *et al.*, 2003) might lead to results that favor the combined spectral and textural models. Also the application of more recent deep-learning concepts as e.g. Convolutional Neural Networks that might offer new ways to include texture and neighborhood information that could lead to improved rainfall estimations. This becomes even more attractive as improved software tools for deep learning are recently evolving that further support high performance computing to significantly reduce computation times.

In summary, we could show that textural variables in optical rainfall retrievals that use simple machine learning architecture and recursive feature selection could not improve the final performance compared to spectral-only models. Therefore, in order to avoid high computation time it is reasonable to retain the pixel-based approach which requires the spectral channels of the optical satellite system as predictors only. A potential reason for the similar performance of the models could rely either in the simple architecture of the neural network or in the spectral resolution of modern sensors which allow the (direct or implicit) retrieval of a variety of cloud optical/geometrical and microphysical properties. For example, studies by e.g. Rosenfield and Fitzpatrick-Lins (1986); Nauss and Kokhanovsky (2006) have shown that large enough combinations of the effective cloud droplet radius and the optical cloud thickness are good predictors for rainfall. In the end, such microphysical insights into precipitation generating processes might be as effective in estimating rainfall area and rates as textural variables which focus on morphometric cloud features.

Acknowledgments This work was supported by the Federal Ministry of Education and Research (BMBF) under Grant number 01LL1301. The authors are grateful to the German Weather Service (DWD) for providing the RADOLAN RW and the EUMETSAT Earth Observation Portal (<https://eoportal.eumetsat.int/>) for providing the Meteosat dataset.

Chapter 4

Satellite-based high-resolution mapping of rainfall over southern Africa

Hanna Meyer (1), Johannes Drönner (2), Thomas Nauß (1)

- (1) Environmental Informatics, Faculty of Geography, Philipps-University Marburg, Deutschhausstr. 10, 35037 Marburg, Germany
- (2) Database Research Group, Faculty of Mathematics und Informatics, Philipps-University Marburg, Hans-Meerwein-Str. 6, 35032 Marburg, Germany

Published in

Atmospheric Measurement Techniques 2017, 10,
2009-2019

Received 01 Feb 2017

Revised 21 Apr 2017

Accepted 09 May 2017

Published 06 Jun 2017

4 Satellite-based high-resolution mapping of rainfall over southern Africa

Hanna Meyer, Johannes Dröner, Thomas Nauss

Abstract

A spatially explicit mapping of rainfall is necessary for Southern Africa for eco-climatological studies or nowcasting but accurate estimates are still a challenging task. This study presents a method to estimate hourly rainfall based on data from the Meteosat Second Generation (MSG) Spinning Enhanced Visible and Infrared Imager (SEVIRI). Rainfall measurements from about 350 weather stations from 2010-2014 served as ground truth for calibration and validation. SEVIRI and weather station data were used to train neural networks that allowed the estimation of rainfall area and rainfall quantities over all times of the day. The results revealed that 60 % of recorded rainfall events were correctly classified by the model (Probability Of Detection, POD). However, the False Alarm Ratio (FAR) was high (0.80), leading to a Heidke Skill Score (HSS) of 0.18. Estimated hourly rainfall quantities were estimated with an average hourly correlation of $\rho = 0.33$ and a Root Mean Square Error (RMSE) of 0.72. The correlation increased with temporal aggregation to 0.52 (daily), 0.67 (weekly) and 0.71 (monthly). The main weakness was the overestimation of rainfall events. The model results were compared to the Integrated Multi-satellitE Retrievals for GPM (IMERG) of the Global Precipitation Measurement (GPM) mission. Despite being a comparably simple approach, the presented MSG based rainfall retrieval outperformed GPM IMERG in terms of rainfall area detection where GPM IMERG had a considerably lower POD. The HSS was not significantly different compared to the MSG based retrieval due to a lower FAR of GPM IMERG. There were no further significant differences between the MSG based retrieval and GPM IMERG in terms of correlation with the observed rainfall quantities. The MSG based retrieval, however, provides rainfall in higher spatial resolution. Though estimate rainfall from satellite data remains challenging especially at high temporal resolutions, this study showed promising results towards improved spatio-temporal estimates of rainfall over Southern Africa.

4.1 Introduction

The dynamics of rainfall play an important role in Southern Africa especially in the arid and semi-arid areas where farming is a main income and the quality of the pastures mainly depends on water availability (Fynn and O'Connor, 2000). Accurate nowcasting of rainfall at high temporal and spatial resolutions is therefore of interest for the farmers in Southern Africa and would help them to assess the carrying capacity of their land. It is of further importance as a baseline product for a variety of environmental research studies as rainfall is a key variable for many ecological and hydrological processes.

Rain gauges are still considered as the most accurate way to measure rainfall. Southern Africa features a network of rain gauges operated by the weather services of the individual countries as well as by a variety of research projects. However, the network does not feature a sufficient density to capture spatially highly variable rainfall dynamics. To obtain spatially explicit data, ground-based radar networks are well established to measure rainfall in other parts of the world (e.g. RADOLAN in Germany, Bartels *et al.*, 2004). A radar network covering the entire region of Southern Africa, however, is currently not available and the existing radar-based rainfall estimates in South Africa are still afflicted with many uncertainties (IPWG, 2016). A satellite-based monitoring of rainfall is therefore an obvious alternative.

A number of global satellite-derived products have been developed in the last decades (e.g. TRMM, CMORPH, PERSIANN, see review in Kidd and Huffman, 2011; Prigent, 2010; Thies and Bendix, 2011; Kidd *et al.*, 2011; Levizzani *et al.*, 2002). Since 2014, the latest product from the Global Precipitation Measurement (GPM) mission, as a successor of the Tropical Rainfall Measuring Mission (TRMM), provides the most recent global estimates of precipitation at high spatial and temporal resolutions. It might be expected that the GPM products would feature a high degree of accuracy since the TRMM-3B42 product has been identified as the most accurate retrieval at least for east Africa (Cattani *et al.*, 2016).

In addition to global rainfall retrievals, a number of regionally adapted retrievals were developed in the last decades (Kühnlein *et al.*, 2014a,b; Meyer *et al.*, 2016b; Feidas and Giannakos, 2012; Giannakos and Feidas, 2013). Kühnlein *et al.* (2014b,a) and Meyer *et al.* (2016b) presented a methodology to estimate rainfall from optical Meteosat Second Generation (MSG) Spinning Enhanced Visible and InfraRed Imager (SEVIRI) data for Germany. In this approach, machine learning algorithms were used to relate the spectral properties of MSG to reliable radar data as a ground truth. Though the retrieval showed promising results, such spatially comprehensive ground truth data are lacking for Southern Africa. An

adaptation of the retrieval technique to Southern Africa hence requires a model training that relies on sparse weather station data as a ground truth.

This study aims to test the suitability of a MSG and artificial neural network based rainfall retrieval which is regionally trained using rain gauge data to provide spatially explicit estimates of rainfall areas and rainfall quantities for Southern Africa. The suitability of the model is assessed by validation with independent weather station data and comparison to the Integrated Multi-satellitE Retrievals for GPM (IMERG) product.

4.2 Methods

The methodology is divided into a pre-processing of satellite and rain gauge data, model tuning and training including its validation, model estimation and comparison to GPM IMERG (Fig. 4.1).

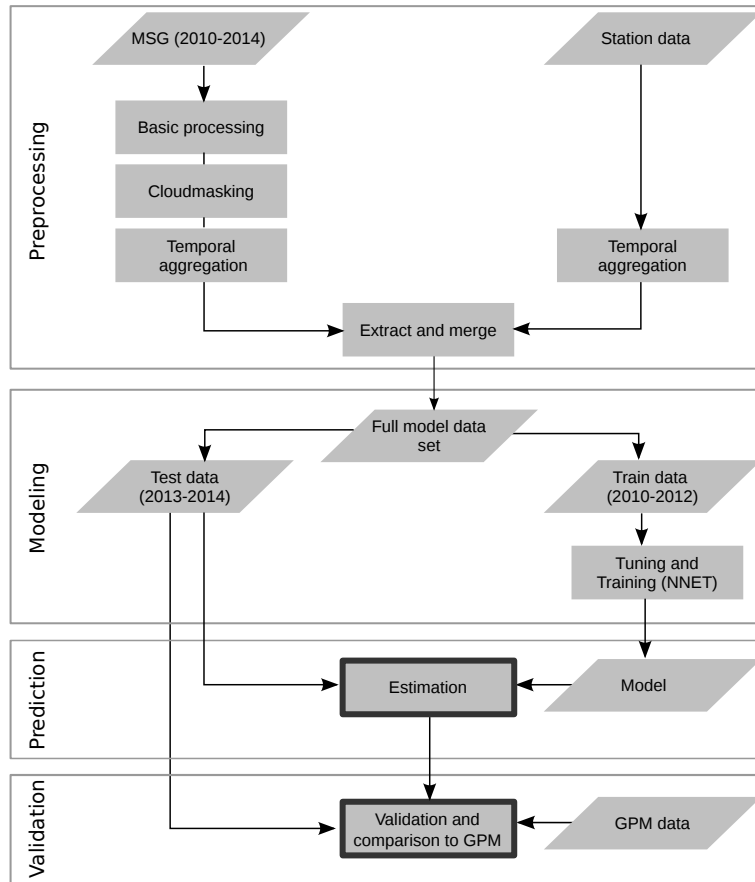


Figure 4.1: Flow chart of the methodology applied in this study.

4.2.1 Study area

The area of investigation comprises South Africa, Lesotho and Swaziland, Namibia, Botswana, Zimbabwe as well as parts of Mozambique (Fig. 4.2). Average annual rainfall in Southern Africa roughly follows an aridity gradient from the dry west to the more humid east. With the exceptions of some coastal regions in South Africa, most rain falls during the summer months. In the coastal areas of South Africa, frontal systems cause light rain that may last over several days. The majority of interior areas are dominated by local and short-term convective heavy showers mostly with thunder in the afternoon or evening hours. Rain from synoptic systems lasting up to several days also occurs. Snow and hail only contribute a negligible amount to the overall precipitation totals. The inter-annual variability of rainfall is high for the arid areas. For a detailed description of Southern African rainfall characteristics see Kruger (2007) and Kaptué *et al.* (2015).

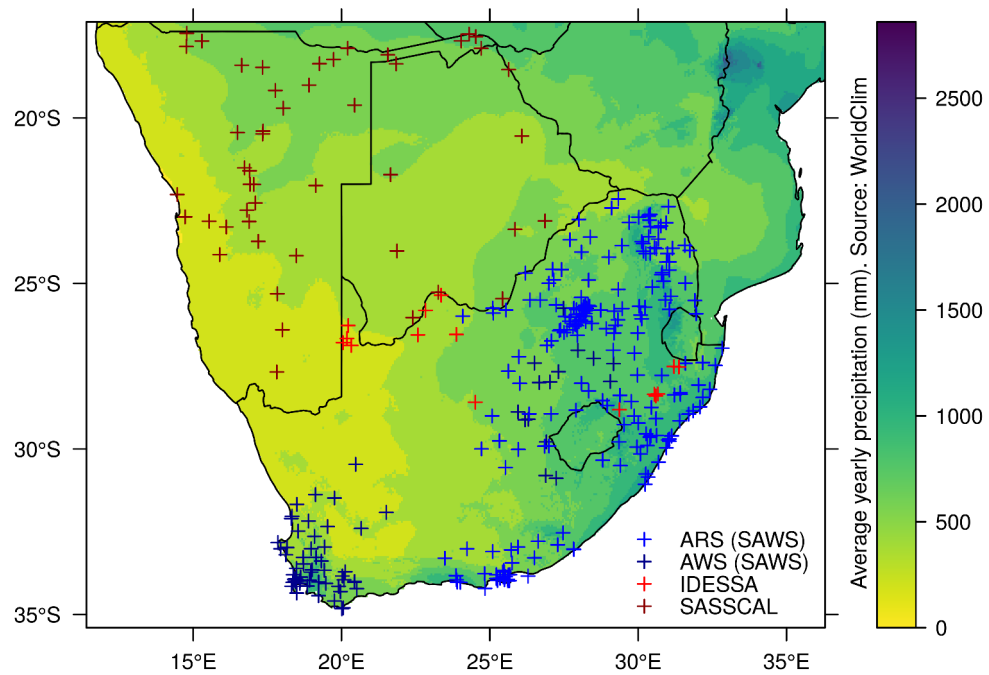


Figure 4.2: Map of the average annual precipitation sums in the study area as estimated by WordClim (Hijmans *et al.*, 2005). Points show the locations of the weather stations that were used as ground truth data in this study. Automatic Rainfall Stations (ARS) and Automatic Weather Stations (AWS) are operated by the South African Weather Service (SAWS). Further stations are operated by SASSCAL WeatherNet as well as by the IDESSA project.

4.2.2 Data and Preprocessing

4.2.2.1 Station data

Rainfall data for 2010 to 2014 were obtained from the South African Weather Service (SAWS). The data were recorded at 229 automatic rainfall stations and 91 automatic weather stations (Fig. 4.2). They were complemented by 22 stations from SASSCAL WeatherNet (www.sasscalweathernet.org) located in southern Namibia and Botswana. For 2014, data from an additional 15 stations in South Africa operated by the IDESSA project (An Integrative Decision Support System for Sustainable Rangeland Management in Southern African Savannas, www.idessa.org) were available. The data passed general provider-dependent quality checks before it was used in this study. This includes filtering of data beyond common data ranges, or situational checks for consistency with related parameters (e.g. air humidity) by SASSCAL. SAWS payed attention to rainfall values > 10 mm within 5 minutes and deleted those values if unreliable. Data from all providers was then included in an on-demand processing database system (Wöllauer *et al.*, 2015) where it was automatically cross-checked for reliability by filtering values < 0 and > 500 mm of rainfall per hour. All station data that provided sub-hourly information was aggregated to a temporal resolution of 1 hour within the database. Though the station data is not randomly distributed in the model domain, it covers the entire aridity gradient, from sites with very low (< 200 mm) precipitation to sites in areas with highest (~ 1500 mm) yearly precipitation sums.

4.2.2.2 Satellite data

MSG SEVIRI (Aminou *et al.*, 1997) scans the full disk every 15 minutes with a spatial resolution of 3×3 km at sub-satellite point (3.5×3.5 km in Southern Africa). Reflected and emitted radiances are measured by 12 channels, three channels at visible (VIS) and very near infrared wavelengths (NIR, between 0.6 and $1.6 \mu\text{m}$), eight channels ranging from near-infrared to thermal infrared wavelengths (IR, between 3.9 and $14 \mu\text{m}$) and one high-resolution VIS channel with a spatial resolution of 1×1 km which was not considered in this study.

The rainfall retrieval technique presented here works under the assumption that VIS, NIR and IR channels of MSG SEVIRI provide proxies for microphysical cloud properties, which are, in turn, related to rainfall. VIS and NIR channels have been shown to be related to cloud optical depth (Roebeling *et al.*, 2006; Benas *et al.*, 2017) and cloud water path (Kühnlein *et al.*, 2014a) where the NIR channel is further related to cloud particle size (Roebeling *et al.*, 2006). The IR channels have been shown to provide information about the cloud top temperature which was used as a proxy for cloud height (Hamann *et al.*, 2014).

The cloud droplet effective radius as well as liquid water path during night was approximated using IR differences (Merk *et al.*, 2011; Kühnlein *et al.*, 2014a).

MSG SEVIRI Level 1.5 data (EUMETSAT, 2010) was preprocessed to radiance values according to EUMETSAT (2012a) and brightness temperatures according to EUMETSAT (2012b) using a processing scheme based on a custom raster processing extension of the eXtensible and fleXible Java library (see <https://github.com/umr-dbs/xxl>) which enables parallel raster processing on CPUs and GPUs using OpenCL.

4.2.2.3 Cloud mask

A cloudmask was used to exclude all pixels that were not cloudy in the respective SEVIRI scenes. For 2010 to 2012, the CM SAF CMA Cloudmask product (Kniffka *et al.*, 2014) was applied. Due to the availability of the CM SAF CMA cloudmask dataset which was currently limited to the years 2004 to 2012, we used the cloud mask information of the CLAAS-2 data record (Finkensieper *et al.*, 2016) for the years 2013 and 2014 which is the 2nd edition of the SEVIRI-based cloud property data record provided by the EUMETSAT Satellite Application Facility on Climate Monitoring (CM SAF; see also Stengel *et al.* (2014) for further information on CLAAS). All pixels that were classified as cloud contaminated or cloud filled were interpreted as cloudy. Pixels that were classified as cloud-free were excluded from further analysis.

4.2.3 Model strategies for rainfall estimation

4.2.3.1 General model framework

The modeling methodology follows the study of Kühnlein *et al.* (2014b,a) who used the spectral channels of MSG SEVIRI to train a Random Forest model that is able to spatially estimate rainfall areas and rainfall rates over Germany. Based on this study, Meyer *et al.* (2016b) have shown that neural networks outperform the initially used Random Forest algorithm. In these previous studies on the rainfall retrieval, the radar based RADOLAN product (Bartels *et al.*, 2004) was used as ground truths to train the model. The high data quality and spatially explicit information allowed the model to be optimised without too much confusion caused by uncertainties in the training data. However, the goal of the retrieval was that it can be applied to areas where spatially explicit data for rainfall are not available, as it is the case in Southern Africa.

4.2.3.2 Training and test data sets

Cloud masked MSG data from 2010 to 2014 were extracted at the locations of the weather stations. To match the temporal resolution of all available rain gauge data, the extracted data were aggregated to hourly values. This was done by taking the median value of the four scenes available every hour. However, only if all four scenes were masked as cloudy, the corresponding hourly values for a respective station were used for further analysis. The extracted and aggregated MSG data were then matched with the corresponding rain gauge information under consideration of the time shift between MSG data (UTC) and rain gauge data (UTC + 2).

The spectral channels as well as the channel differences $\Delta T6.2 - 10.8$, $\Delta T7.3 - 12.1$, $\Delta T8.7 - 10.8$, $\Delta T10.8 - 12.1$, $\Delta T3.9 - 7.3$, $\Delta T3.9 - 10.8$ and the sun zenith were used as predictor variables during daytime, in accordance to (Kühnlein *et al.*, 2014a) and previous studies on MSG based delineation of cloud properties (see section 4.2.2.2). Thus, the predictor variables contain the SEVIRI channels as well as channel combinations. Although this partially duplicates information, the channel combinations allow highlighting patterns that might not be apparent in the individual channels. As additional potential predictors, Meyer *et al.* (2017b) tested different cloud texture parameters and have shown that the chosen spectral channels and differences are sufficient as predictors.

Since neural networks require that the predictor variables are standardized, all predictors were centered and scaled by dividing the values of the mean-centered variables by their standard deviations. Since the VIS and NIR channels of MSG are not available during the nighttime, the dataset was split into a daytime dataset (data points with a solar zenith angle $< 70^\circ$) and a nighttime dataset (data points with a solar zenith angle $> 70^\circ$) and were considered in separate models. Though two different models might lead to rough transitions between daytime and nighttime estimates, accurate estimates were in the foreground of this study, leading to the decision of separate models according to data availability. The response variables (rainfall yes/no and rainfall quantities) were taken from the rain gauge measurements.

The years 2010 to 2012 were used for model training. The year 2013 was used for validation. The retrieval process was two-step and consisted of (i) the identification of precipitating cloud areas and (ii) the assignment of rainfall quantities. All 2010 to 2012 data from the rain gauges that are masked as cloudy by the cloud mask products were used for training the rainfall area model. All recorded rainfall events were used for training the rainfall quantities model. The resulting training dataset comprised 917774 (daytime) and 1409072 (nighttime) samples for the rainfall area training and 69703 (daytime) and 129325 (nighttime) samples for training of rainfall quantities from 26243 individual MSG scenes.

4.2.3.3 Tuning and model training

A single-hidden-layer feed-forward neural network was applied as machine learning algorithm. The spectral channels of MSG SEVIRI as well as the channel differences served as input nodes (predictor variables). The neural network was then applied to learn the relations between these spectral information and rainfall areas or rainfall quantities, respectively. In this context, a sophisticated pre-selection of input variables is not required, as the network is able to deal with correlated and even uninformative predictors unless their number is very high (Meyer *et al.*, 2017b), which was not the case in this study. For the technical realisation, all steps of model training were performed using the R environment for statistical computing (R Core Team, 2016). The neural network implementation from the "nnet" package (Venables and Ripley, 2002; Ripley and Venables, 2016) in R was used in conjunction with the "caret" package (Kuhn, 2016b) that provides enhanced functionalities for model training, estimation and validation.

Neural networks require two hyperparameters to be tuned to avoid under- or overfitting of the data: the number of neurons in the hidden layer, as well as the weight decay. The neurons in the hidden layer represent nonlinear combinations of the input data and their number influences the performance of the model (Panchal *et al.*, 2011). Weight decay penalizes large weights and controls the generalisation of the outcome (Krogh and Hertz, 1992). The number of neurons as well as weight decay were tuned using a stratified 10-fold cross-validation. Thus, the training samples were randomly partitioned into 10 equally sized folds with respect to the distribution of the response variable (i.e., raining cloud pixels, rainfall rate). Thus, every fold is a subset (1/10) of the training samples and has the same distribution of the response variable as the total set of training samples. Models were then fitted by repeatedly leaving out one of the folds. The performance of a model was then determined by predicting on the held back fold. The performance metrics from the held back iterations were averaged to the overall model performance for the respective set of tuning values. For the rainfall areas classification models, the distance to a "perfect model", based on Receiver Operating Characteristics (ROC) analysis (see Meyer *et al.* (2016b) for its application in rainfall retrievals) was used as decisive performance metric. For the rainfall quantities regression models, the Root Mean Square Error (RMSE) was used. The number of hidden units were tuned for each value between two and the number of predictor variables. Weight decay was tuned between 0 and 0.1 with increments of 0.02 (Kuhn and Johnson, 2013b). For training of rainfall areas, the threshold that separates rainy from non-rainy clouds according to the estimated probabilities was an additional tuning parameter. The optimal threshold was expected to be considerably smaller than 0.5 since the amount of non rainy samples was higher than the amount of rainy samples. Therefore, the range of tested thresholds was 0 to 0.1 with increments of 0.01, and 0.4 to 1 with

increments of 0.1. See Meyer *et al.* (2016b) for further details of the threshold tuning methodology.

The optimal values for the hyperparameters that were revealed in the tuning study (Tab. 4.1) were adopted for the final model fitting. In this step, the model is fit to all training data using the optimal hyperparameters.

Table 4.1: *Optimal hyperparameters for the individual models revealed during the tuning study and applied in the final model fitting.*

	No. of neurons	Weight decay	Threshold
Rainfall areas at daytime	5	0.05	0.07
Rainfall areas at nighttime	5	0.07	0.01
Rainfall quantities at daytime	5	0.05	
Rainfall quantities at nighttime	5	0.05	

4.2.3.4 Spatial estimations of rainfall

Final models were applied to all hourly MSG SEVIRI scenes from 2010-2014 for the Southern Africa extent to obtain spatio-temporal estimates of rainfall. Therefore, the clouded areas of a scene were first classified into rainy or not rainy using the respective model. The rainfall quantities were then estimated for the estimated rainfall areas. To ensure consistency within one scene, the choice of the model being applied (either the daytime or nighttime model) was made according to the mean solar zenith angle of the respective scene. If the mean solar zenith angle was $< 70^\circ$, rainfall for the entire scene was estimated using the daytime model. For scenes with a mean solar zenith angle $> 70^\circ$, the nighttime model was applied.

4.2.4 Validation

Model estimates and weather station records from the entire year 2013 were used as independent data for model validation. For the validation of estimated rainfall areas, all pixels at the location of the weather stations that were classified as cloudy by the cloud mask product were considered. Therefore the information from the weather stations about whether it was raining or not was compared to the model estimate for the respective MSG pixel. The validation data contained 403211 samples during daytime and 565415 samples during nighttime. Average hourly Probability Of Detection (POD), Probability Of False Detection (POFD), False Alarm Ratio (FAR) and Heidke Skill Score (HSS) were calculated as validation metrics. The POD gives the percentage of rain pixels that the model correctly identified as rain (Tab. 4.2, 4.3). POFD gives the proportion of non-rain pixels that the model incorrectly classified as rain. The FAR gives the proportion

of estimated rain where no rain is observed. The HSS also accounts for chance agreement and gives the proportion of correct classifications (both rain pixels and non-rain pixels) after eliminating expected chance agreement.

Table 4.2: *Confusion matrix as baseline for the calculation of the verification scores used for the validation of the rainfall area estimates.*

Estimation	Observation		
		Rainfall	No Rainfall
	Rainfall	True positives (TP)	False positives (FP)
	No Rainfall	False negatives (FN)	True negatives (TN)

Table 4.3: *Categorical metrics for validation of rainfall area estimates.*

Metric	Formula	Range	optimal value
Probability Of Detection	$POD = \frac{TP}{TP+FN}$	0 - 1	1
Probability Of False Detection	$POFD = \frac{FP}{FP+TN}$	0 - 1	0
False Alarm Ratio	$FAR = \frac{FP}{TP+FP}$	0 - 1	0
Heidke Skill Score	$HSS = \frac{TP*TN-FP*FN}{[(TP+FN)*(FN+TN)+(TP+FP)*(FP+TN)]/2}$	$-\infty - 1$	1

To evaluate the ability of the model to estimate rainfall quantities, the correlation between the measured and the estimated hourly rainfall was calculated using Spearman’s Product Moment Correlation (ρ) to account for a non-normal distribution of the data. RMSE was also calculated. All cloudy data points (including non-rainy data points) were used for the validation of rainfall quantities. The rainfall quantities were further aggregated to daily, weekly and monthly rainfall sums to assess the performance of the model on different temporal scales.

4.2.5 Comparison to GPM

The results of the presented rainfall retrieval were compared to the rainfall estimates of the GPM mission. GPM, as a successor of the Tropical Rainfall Measuring Mission (TRMM), consists of an international network of satellites designed for worldwide high resolution precipitation estimates (Hou *et al.*, 2014; Skofronick-Jackson *et al.*, 2017). GPM provides data from March 2014 onwards. The GPM IMERG product estimates rainfall by combining all available passive-microwave estimates as well as microwave-calibrated infrared satellite estimates and data from rainfall gauges. GPM IMERG is available in 6h, 18h and 4 months latency.

In this study the 4 month latency (final product) with 30 minutes temporal and 0.1° spatial resolution ($\sim 10\text{km} \times 10\text{km}$) was used (Huffman *et al.*, 2014). Due to different data availabilities of GPM IMERG, MSG as well as weather station data, the comparison was conducted for the overlapping time period late March 2014 to August 2014. GPM was aggregated from 30 minutes to 1h to match the temporal resolution of the MSG based estimates. Both products were validated

using the weather station data as a reference. The performance metrics were compared between the MSG product and the GPM product on an hourly basis.

4.3 Results

4.3.1 Model performance

On average, 60 % of the rainfall observations were correctly identified as rainy by the model with a high number of scenes having much higher PODs (Fig. 4.3). The POFD was low (18 % in average) but the estimates featured a high FAR of 0.80. The average HSS per scene was 0.18. The POD was highest for high measured rainfall quantities and decreased for lower rainfall quantities (Fig. 4.4). FAR was highest for low predicted rainfall quantities and decreased for higher predicted quantities.

The average hourly RMSE was 0.72 mm h^{-1} (Fig. 4.5). Especially data points with low or medium measured rainfall could be estimated with low RMSE (Fig. 4.4). The RMSE was higher for high measured rainfall. Correlation indicated by Spearman's rho was 0.33 on hourly average. The performance of modeled rainfall quantities increased with the aggregation level (Fig. 4.6). The average correlation increased from $\rho = 0.33$ (hourly) to 0.52 on a daily, 0.67 on a weekly and 0.71 on a monthly basis. An overestimation of rainfall is observed especially when aggregated to monthly totals. An example of temporally aggregated rainfall estimates for 2013 are shown in Fig. 4.7.

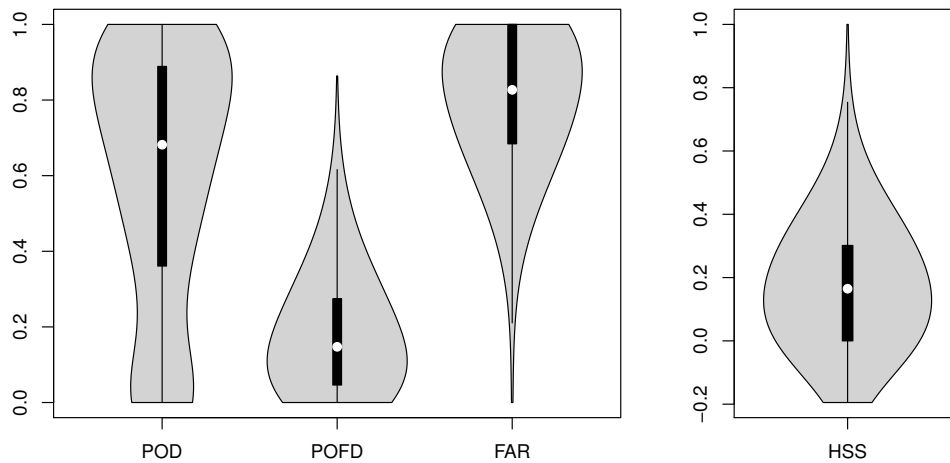


Figure 4.3: Validation of estimated rainfall areas for 2013 on an hourly basis. Each of the data points is the average performance of one hour. The data are visualized as "vioplot" where a boxplot is complemented by the kernel density of the data shown as grey areas at the sides of the boxplot.

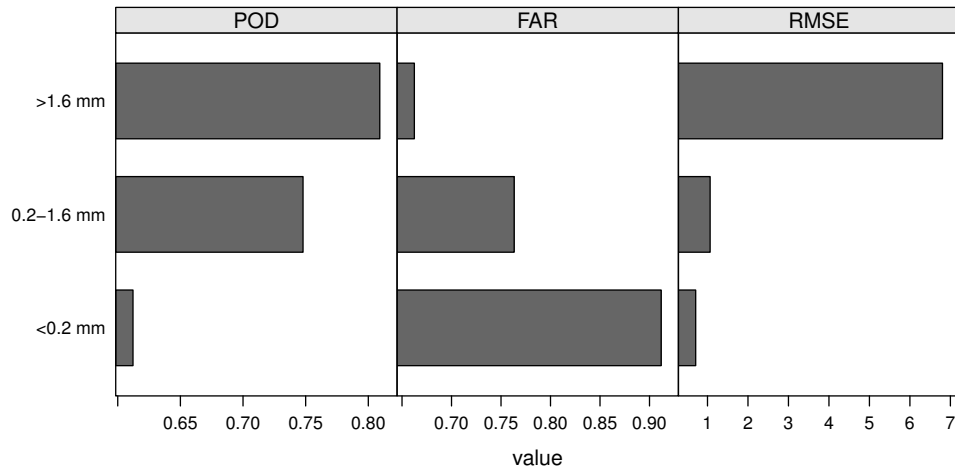


Figure 4.4: Comparison of POD for different hourly measured rainfall quantities as well as FAR for different predicted rainfall quantities. RMSE was compared for different measured rainfall quantities. All data points from 2013 were used for the calculation of the statistics. Thresholds for the three rainfall classes were set according to the first and third quartiles of the measured hourly rainfall quantities.

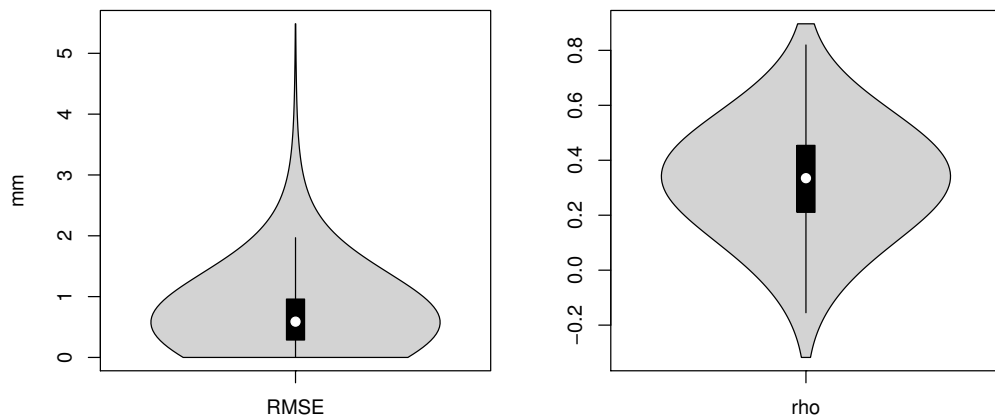


Figure 4.5: Validation of estimated rainfall quantities for 2013 on an hourly basis. Each of the data points is the average performance of one hour. See Fig. 4.3 for further information on the figure style.

4.3.2 Comparison to GPM

Compared to GPM IMERG, the MSG based rainfall retrieval for the period Mar-Aug 2014 showed a higher POD (0.57) than GPM IMERG (0.28) which considerably underestimated rainfall events (Fig. 4.8). In contrast, GPM IMERG had a lower FAR (0.70) than the MSG based model (0.81). However, the FAR was high for both retrievals. The average HSS was the same for both retrievals (0.17), but the median HSS for GPM IMERG was 0 which was considerably lower

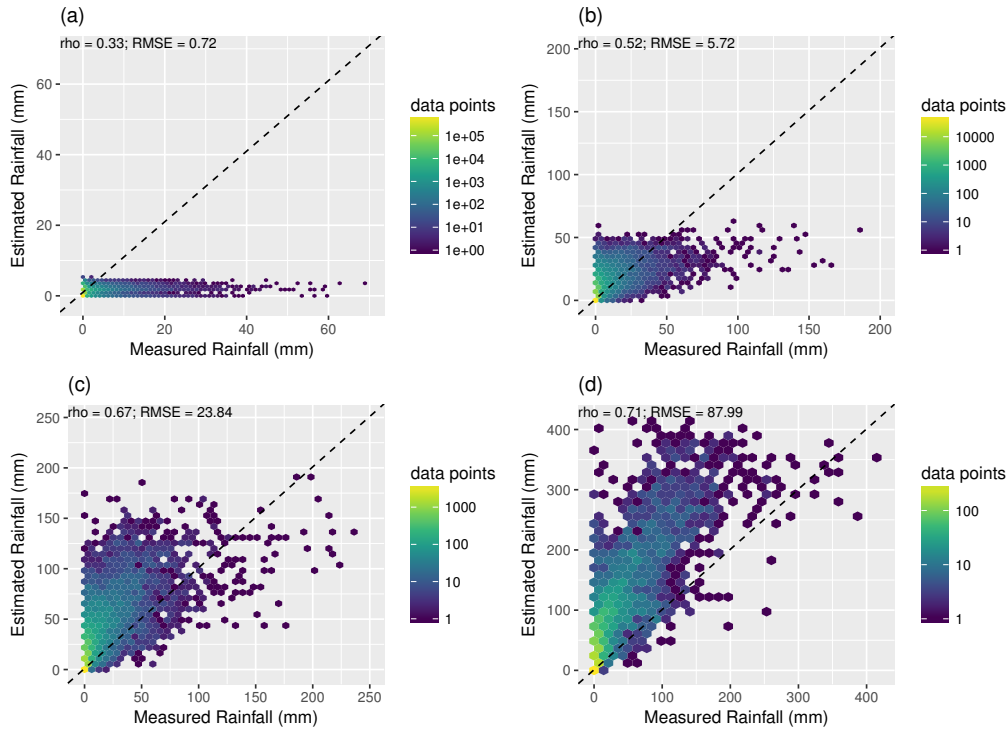


Figure 4.6: Validation of estimated rainfall quantities for 2013 at (a) hourly resolution and on the different aggregation (b) daily, (c) weekly, (d) monthly. Each of the data points represents a station at the respective level of temporal aggregation. Rho represents the average correlation for each time step of the respective aggregation level. For an easy visual interpretation, the data are presented via hexagon binning where the number of data points falling in each hexagon are depicted by color.

than using the MSG based retrieval (0.10). Concerning the rainfall quantities, neither the correlation to measured rainfall nor the RMSE showed significant differences between both retrievals (Fig. 4.9). The average rho was 0.36 for the MSG based retrieval and 0.34 for GPM IMERG. The average RMSE was 0.88 for the MSG based retrieval and 0.85 for MSG IMERG.

Fig. 4.10 gives an example of the differences between the MSG based retrieval and GPM IMERG for 2014/04/24 12:00 UTC where severe floods occurred in the Eastern Cape province of South Africa. The colour composite of the corresponding MSG scene shows that clouds had a high optical depth in this area. The pattern is reflected in the estimates of the MSG based retrieval that estimated rainfall for the areas with high values of optical depth. This was partly confirmed by the weather station data. However, rainfall was also estimated for areas where weather stations did not record any rainfall. In contrast, GPM IMERG showed an underestimation of rainfall areas, but still captured the high rainfall quantities that were recorded by the weather stations. The summary statistics for this hour are a POD of 0.75 for the MSG based retrieval and 0.19 for GPM IMERG. FAR was 0.65 and HSS 0.34 for the MSG based retrieval compared to a FAR

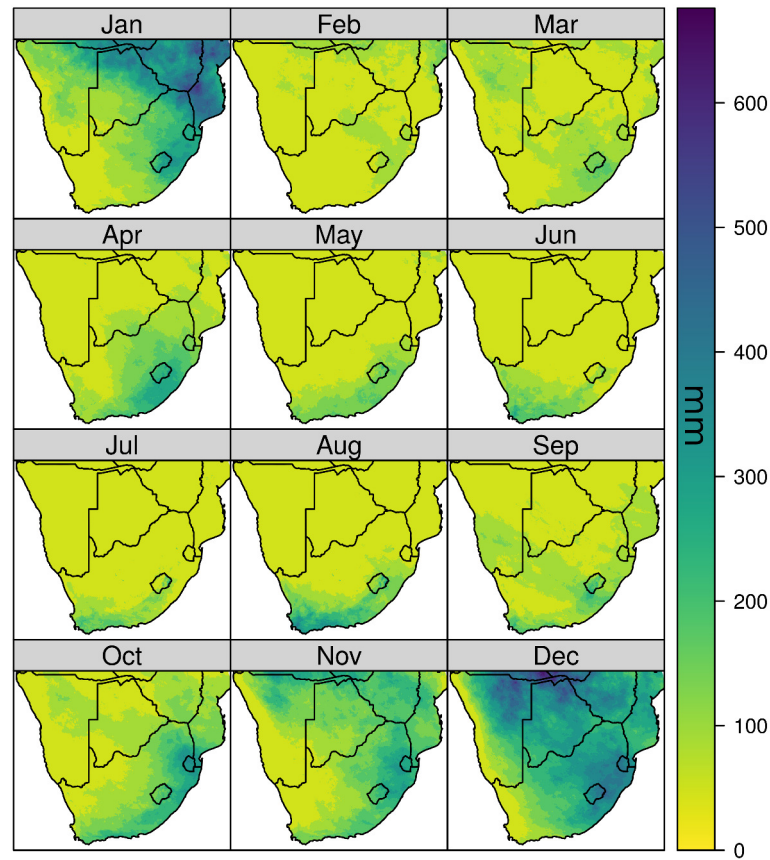


Figure 4.7: *Monthly precipitation sums in mm of the year 2013 as estimated by this study.*

of 0.89 and a HSS of 0.08 for GPM IMERG. The correlation between estimated and observed rainfall was 0.39 for the MSG based retrieval and -0.06 for GPM IMERG.

4.4 Discussion

The presented monthly maps reflect the general spatial and temporal rainfall patterns of Southern Africa as shown in Kruger (2007). They also reflect the annual characteristics of the year 2013. For example, the heavy rainfall events over southern Mozambique and the Limpopo River basin during mid January (Manhique *et al.*, 2015).

The validation of the rainfall retrievals showed promising results but also highlights the difficulties of optical satellite-based rainfall estimates. The strength of the retrieval in terms of rainfall areas classification was a high POD for heavy rainfall events. The rainfall quantities for the heavy rainfall events were, how-

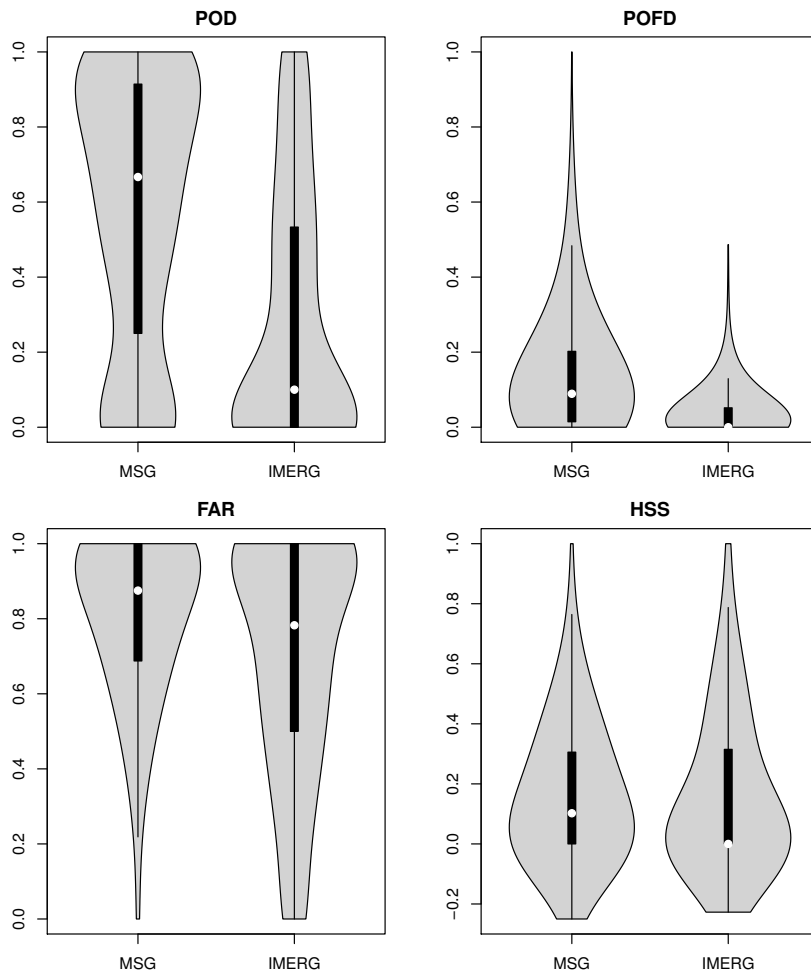


Figure 4.8: Comparison of the performance of the MSG based retrieval and GPM IMERG for rainfall area delineation between March and August 2014. Each of the data points is the average performance of one hour. See Fig. 4.3 for further information on the figure style.

ever, underestimated in most cases. The major problem of the model was the overestimation of rainfall events leading to an overestimation of rainfall quantities. However, false alarms in the retrieval were generally predicted with low rainfall quantities. In this context, it is of note that in view to the scene-based validation strategy, FAR can easily increase in dry conditions when there are just a few false alarms in the estimates and no rainfall was observed by any station. However, the FAR was still high for hours with a considerable number of rainfall events. This might be partly explainable by spatial displacement due to parallax shifts. Though the shift is generally below 1 pixel in this region, even minor shifts can affect model training as well as the estimates. For future enhancement of the rainfall retrieval, a correction of the parallax shift (Vicente *et al.*, 2002) would be appropriate. Differences in spatial and temporal scale are also an important issue especially since a majority of rainfall events in Southern Africa are of small spatial and temporal extent. The aggregation to an hour as well as the

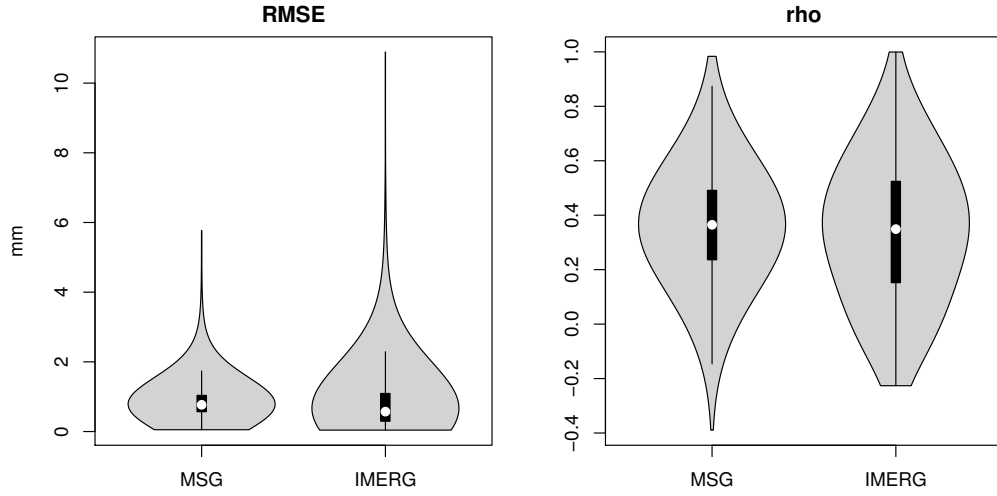


Figure 4.9: Comparison of the performance of the MSG based retrieval and GPM IMERG for hourly rainfall quantities between March and August 2014. Each of the data points is the average performance of one hour. See Fig. 4.3 for further information on the figure style.

assumption that the weather station observation is representative for the entire pixel are also problematic, though essential. The issue of scale especially affects the broader resolution GPM IMERG data where a several km sized pixel is validated by a single point measurement. Beside of the issue of scale and spatial displacement, the retrieval technique depends on the quality of the rain gauge observations. Although the data was quality checked, common problems associated with rain gauge measurements e.g. wind drift or evaporation leading to errors in the ground truth data and affect model training and validation remain (Kidd and Huffman, 2011). Also, due to different installation dates of the individual weather stations as well as the natural challenge of maintaining weather stations in remote areas, no gapless dataset could be compiled. Therefore, different sensor and data provider dependent calibration techniques, gaps in the time series of the data as well as the general problems associated with rain gauge measurements might lead to inconsistencies and uncertainties. However, no reliable alternatives are available and rain gauge measurements are still considered as most reliable source of rainfall data.

The retrieval techniques relied on the cloud mask for an initial selection of relevant data points used for model training, validation and the final spatio-temporal estimates. Therefore, it can't be excluded that some data points were falsely excluded from the analysis as they were falsely masked as being not cloudy but rainfall was measured on the ground. However, we assume that rainy clouds are easy to capture by common cloud masking algorithms and that the resulting bias is therefore comparably small.

Despite the errors and uncertainties associated with the presented rainfall

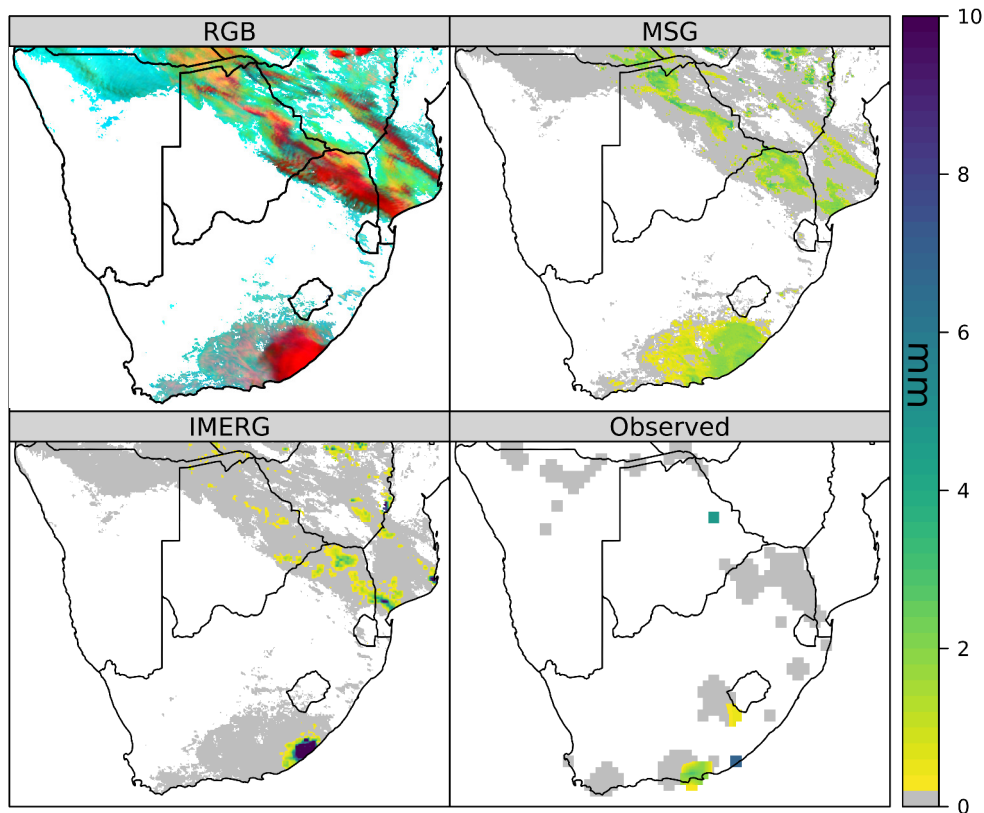


Figure 4.10: Sample satellite scene from 2014/04/24 10:00 UTC represented as a VIS0.8-IR3.9-IR10.8 false colour composite according to (Rosenfeld and Lensky, 1998) where cloud optical depth is indicated by red colouration, cloud particle sizes and phases in green and the brightness temperature modulates in blue. The rainfall estimates for this scene (estimated using the daytime model) are shown as well as the corresponding GPM IMERG product. Observed rainfall is depicted where weather station data were available. For visualization purposes, the spatial extent of the stations was increased. White background in the colour composite as well as in the MSG based retrieval and the GPM IMERG product represent no data due to missing clouds. In addition, white background in the representation of the observed rainfall is due to the absence of weather stations.

retrieval, the combination of MSG data and neural networks are a promising approach. The model presented in this study outperformed the GPM IMERG product in terms of rainfall area detection where GPM IMERG considerably underestimated rainfall events. This behavior is partly explainable by scale because GPM IMERG has a coarser resolution of 0.1° . This makes local processes difficult to capture which is an disadvantage considering that in Southern Africa especially small scale convective showers contribute to rainfall sums Kruger (2007). In terms of rainfall quantities, GPM IMERG and the presented retrieval did not show significant differences in correlation. The sample spatial comparison has shown that GPM IMERG has more differentiated rainfall estimates while the MSG based retrieval tends to estimate the mean distribution.

The presented MSG based retrieval is an easy to use method and allows for time series at a relatively high spatial resolution. Aside of the promising results compared to GPM IMERG, the daily estimates of the MSG based retrieval are at least comparable to other products incorporated in the IPWG validation study IPWG (2016). A detailed comparison could currently not be given since validation data and strategy were not identical. Incorporation of the presented retrieval scheme to the IPWG validation study is intended by the authors for future assessment.

4.5 Conclusions

The rainfall retrieval technique developed in this study provides hourly rainfall estimates at high spatial resolution based on the spectral properties of MSG SEVIRI data and neural networks. The retrieval showed promising results in terms of rainfall area detection and estimation of rainfall quantities. However, the results also showed that the estimation of rainfall remains challenging. The main weakness of the presented retrieval was the overestimation of rainfall occurrence. However, the retrieval could compete with the GPM IMERG product in terms of rainfall quantity and was even better for rainfall area detection.

High resolution spatial datasets of rainfall is requested by a variety of research disciplines. The developed MSG based rainfall retrieval is able to deliver time series from the launch of MSG SEVIRI onward. An operationalization for near real-time rainfall estimates is intended. It can therefore serve as valuable dataset where high resolution rainfall for Southern Africa are needed. As an example it will serve as an important parameter within the "IDESSA" (An Integrative Decision Support System for Sustainable Rangeland Management in Southern African Savannas) project that aims to implement an integrative monitoring and decision-support system for the sustainable management of different savanna types. The hourly and aggregated rainfall quantity estimations are available from the authors on request.

Acknowledgments This work was financially supported by the Federal Ministry of Education and Research (BMBF) within the IDESSA project (grant no. 01LL1301) which is part of the SPACES-program (Science Partnership for the Assessment of Complex Earth System processes). We are grateful to the South African Weather Service for providing us with their rainfall data for South Africa and to SASSCAL WeatherNet for rainfall data from Namibia and Botswana. The cloud masking was done by using Level-2 data of the CLAAS-2 data record provided by the EUMETSATs Satellite Application Facility on Climate Monitoring (CM SAF). The GPM IMERG V3 data were provided by the NASA/Goddard Space Flight Center's Mesoscale Atmospheric Processes Laboratory and Precipitation Processing System (PPS), which develop and compute the GPM IMERG V3 as a contribution to project GPM, and archived at the NASA GES DISC.

Chapter 5

Mapping daily air temperature for Antarctica based on MODIS LST

Hanna Meyer (1), Marwan Katurji (2), Tim Appelhans (1),
Markus U. Müller (3), Thomas Nauss (1), Pierre Roudier (4,5),
Peyman Zawar-Reza (2)

- (1) Environmental Informatics, Faculty of Geography, Philipps-University Marburg, Deutschhausstr. 10, 35037 Marburg, Germany
- (2) Center for Atmospheric Research, University of Canterbury, Christchurch 8020, New Zealand
- (3) Environmental Informatics, Landcare Research, Hamilton 3240, New Zealand
- (4) Soils & Landscapes, Landcare Research, Palmerston North 4442, New Zealand
- (5) Te Punaha Matatini, A New Zealand Centre of Research Excellence, Private Bag 92019, Auckland, 1142, New Zealand

Published in

Remote Sensing, 2016, 8(9), 732

Received 18 May 2016

Revised 4 August 2016

Accepted 31 August 2016

Published 5 September 2016

5 Mapping daily air temperature for Antarctica based on MODIS LST

Hanna Meyer, Marwan Katurji, Tim Appelhans, Markus U. Müller, Thomas Nauß, Pierre Roudier, Peyman Zawar-Reza

Abstract

Spatial predictions of near-surface air temperature (T_{air}) in Antarctica are required as baseline information for a variety of research disciplines. Since the network of weather stations in Antarctica is sparse, remote sensing methods have large potential due to their capabilities and accessibility. Based on the MODIS land surface temperature (LST) data, T_{air} at the exact time of satellite overpass was modelled at a spatial resolution of 1 km using data from 32 weather stations. The performance of a simple linear regression model to predict T_{air} from LST was compared to the performance of three machine learning algorithms: Random Forest (RF), generalized boosted regression models (GBM) and Cubist. In addition to LST, auxiliary predictor variables were tested in these models. Their relevance was evaluated by a Cubist-based forward feature selection in conjunction with leave-one-station-out cross-validation to reduce the impact of spatial overfitting. GBM performed best to predict T_{air} using LST and the month of the year as predictor variables. Using the trained model, T_{air} could be estimated with a leave-one-station-out cross-validated R^2 of 0.71 and a RMSE of 10.51 °C. However, the machine learning approaches only slightly outperformed the simple linear estimation of T_{air} from LST (R^2 of 0.64, RMSE of 11.02 °C). Using the trained model allowed creating time series of T_{air} over Antarctica for 2013. Extending the training data by including more years will allow developing time series of T_{air} from 2000 on.

Keywords Air temperature; Antarctica; Feature selection; Machine learning; MODIS LST

5.1 Introduction

Near-surface air temperature (T_{air}) plays an important role in ecological, glaciological and climatological processes in Antarctica. Climate change further raises the need to study the spatio-temporal trends in T_{air} and its induced regional feedback processes. Therefore, spatially explicit T_{air} datasets are of high interest for the scientific community and research effort for its construction is in high demand (Schneider and Reusch, 2016).

A common approach to obtain spatially explicit T_{air} datasets is spatial interpolation based on station records (Appelhans *et al.*, 2015; Hofstra *et al.*, 2008; Jarvis and Stuart, 2001; Stahl *et al.*, 2006). However, interpolation methods rely on a sufficiently dense number of points. Due to the remoteness of Antarctica, the network of weather stations is sparse (Lazzara *et al.*, 2012) which makes simple interpolation approaches difficult. The applications of these methods to Antarctica are therefore limited to a low temporal resolution of e.g. annual means (Wang and Hou, 2009), rather than aiming at daily products. In order to obtain medium resolution datasets of T_{air} , remote sensing is a promising alternative: it offers spatially explicit proxies for T_{air} , and is therefore suitable for areas with low weather station density, such as Antarctica Rhee and Im (2014). Though T_{air} cannot be directly measured from space, land surface temperature (LST) is a widely used derived product from infrared bands and a proxy for T_{air} Gallo *et al.* (2011) due to surface-atmosphere energy exchange processes.

The Moderate Resolution Imaging Spectroradiometer (MODIS) sensor on-board the Terra and Aqua spacecrafts acquires LST data four times per day (two during the day, two during the night) with a spatial resolution of 1km. MODIS LST was successfully used to estimate T_{air} for various regions of the world (Vancutsem *et al.*, 2010; Colombi *et al.*, 2007; Zhu *et al.*, 2013; Sohrabinia *et al.*, 2014; Mostovoy *et al.*, 2006; Benali *et al.*, 2012; Neteler, 2010; Huang *et al.*, 2015). Kilibarda *et al.* (2014) performed a kriging based spatio-temporal interpolation of global daily temperatures including MODIS LST as predictor. However, since the focus of this study was on a global T_{air} prediction, Antarctica was only marginally represented in the training data. Applications of MODIS LST to predict T_{air} for Antarctica are limited to the study of Wang *et al.* (2013) who compared monthly averages of MODIS LST with T_{air} in the Lambert Glacier Basin in East Antarctica.

The majority of the studies rely on linear regressions or simple bias corrections to estimate T_{air} from LST alone. However, the linearity of the relation is questionable. Colombi *et al.* (2007) found differences in the performance of the linear model according to daytime and altitude. Vancutsem *et al.* (2010) noticed variations in the performance depending on region and season, while Benali

et al. (2012) used several predictors and tested different model structures which included auxiliary predictors. Also Xu *et al.* (2012) included land cover type and altitude in their models to improve the simple linear model between LST and T_{air} . Most recently, Janatian *et al.* (2017) tested the importance of 11 auxiliary variables in addition to LST in a stepwise regression analysis and revealed julian day, altitude and solar zenith angle as effective additional predictors for T_{air} . Emamifar *et al.* (2013) used M5 regression trees to model daily T_{air} in Iran from MODIS LST and auxiliary variables and highlighted the advantage of tree-based models for an operational monitoring of T_{air} . Surprisingly, machine learning algorithms were only rarely applied in the context of estimating T_{air} using LST (Xu *et al.*, 2014; Emamifar *et al.*, 2013). However, those methods are good contenders to model T_{air} , since they can handle non-linearity and highly correlated predictor variables (Kuhn and Johnson, 2013a; James *et al.*, 2013).

The aim of this study is to create a medium resolution spatially explicit daily T_{air} product for Antarctica. In this context, we have tested the performance of different machine learning algorithms to estimate T_{air} from LST and auxiliary variables as an alternative to a simple linear approach.

5.2 Methods

5.2.1 Data and Preprocessing

5.2.1.1 LST

The daily LST data (version 5, Wan, 2008) based on the MODIS sensor onboard the Aqua and Terra satellites are distributed as the MOD11A1 and MYD11A1 products Land Processes Distributed Active Archive Center (LP DAAC) (2013). The MODIS LST products consist of daytime and nighttime measurements at 1 km resolution. Their calculation is based on a split-window algorithm that uses the emissivities from MODIS bands 31 and 32 that were, in turn, calculated using information about land cover type, atmospheric column water vapour and lower boundary air surface Wan (2008). The data are cloud-masked on the basis of the MODIS Cloud Mask algorithm Ackerman *et al.* (1998) that applies typical thresholds in the visible and infrared channels. The MODIS LST data are stated to be very accurate with a deviation of mostly below 1K in the range between -10 °C and 58 °C Wan (2008). Regarding colder environments, Westermann *et al.* Westermann *et al.* (2012) validated MODIS LST for Svalbard in Norway and found a bias of 3K. No decrease in performance could be observed with decreasing temperatures of down to about -40 °C.

In this study, the products were not used as temporal aggregates (e.g., 8-day

composites) but the instantaneous LST values at the respective time of satellite overpass were used. Aqua and Terra pass Antarctica several times per day since the overlap of the orbits increases the closer you get to the poles. The product consists of data from different overpass times. The overpass times of the corresponding LST values are used for each pixel in full hours.

5.2.1.2 Station Records

Three sources of automatic weather station data were used as ground truth in this study. The Antarctic Meteorological Research Center (AMRC) at the University of Wisconsin (Lazzara *et al.*, 2012) provides data from weather stations distributed over the entire continent. Air temperature was measured 3 m above ground level. The Long Term Ecological Research (LTER) programme (Doran *et al.*, 1995) provides weather station data from the McMurdo Dry Valleys where air temperature was measured 3 m above ground level. With focus on soil climate, the United States Department of Agriculture (USDA) provides data from weather stations in the Ross Sea Region (Seybold *et al.*, 2009). Air temperature from the USDA sites was measured at 1.6 m above ground. Temperature sensors of all providers were mounted within radiation shields.

In total, 32 weather stations were used for model training and validation (Figure 5.1). All weather stations provide data in 15 minutes to hourly temporal resolution and were, if necessary, aggregated to one hour. Therefore, all measured T_{air} values that were recorded within each hour were averaged.

5.2.1.3 Auxiliary Data

The Radarsat Antarctic Mapping Project (RAMP) Digital Elevation Model (DEM), version 2 (Liu *et al.*, 2015), was used as one of the auxiliary predictor variables. The 200 m resolution DEM was bilinearly resampled to 1000 m to match the resolution of the MODIS LST data. Slope, aspect, and skyview factor (which describes the fraction of visible sky) were derived from this DEM using SAGA GIS (Conrad *et al.*, 2015). Aspect was classified into north, east, south and west and used as a categorical variable in the model. The Bedmap2 data (Fretwell *et al.*, 2013) were used to classify the landscape into ice covered or ice free areas according to their ice surface elevation information. The month of the year (Jan-Dec), the season (Spring, Summer, Autumn, Winter) and time of day (1–24 h) were included as categorical variables, as well as the sensor type (either Terra or Aqua) to account for potential sensor specific differences in LST.

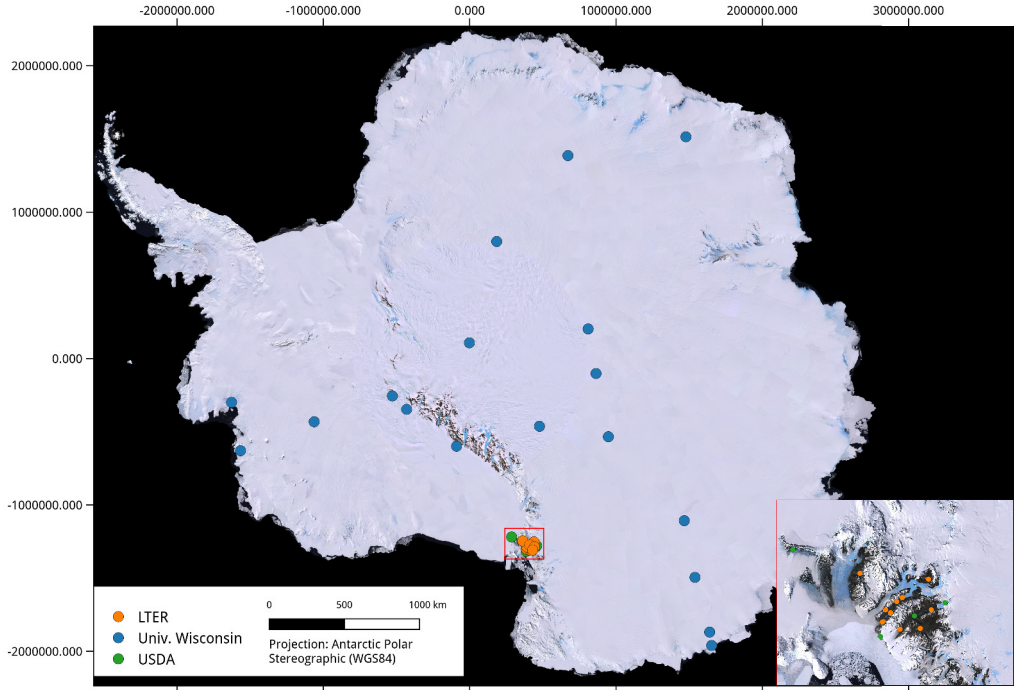


Figure 5.1: Map of weather stations used for model training and evaluation overlaid on a Landsat composite image (U.S Geological Survey, 2007). The McMurdo Dry Valleys are shown in detail.

5.2.1.4 Compilation of Model Training and Testing Data

MODIS LST data, as well as the auxiliary variables were extracted at the location of the weather stations. The MODIS LST values at the respective overpass times were matched with the corresponding station records. For model training, a subset of 40% of the data was used, corresponding to 12280 data points. They were selected by stratified random sampling with respect to the station. The remaining 60% were used as test subset to assess the model performance and the problem of overfitting which will be explained in Section 5.2.2.2.

5.2.2 Modelling

5.2.2.1 Algorithms

A simple linear regression between LST and T_{air} was considered as a baseline model since it is the most intuitive and widely used method to estimate T_{air} from MODIS LST. The three machine learning algorithms Random Forest (RF), generalized boosted regression models (GBM), and Cubist were considered as alternative models. These algorithms were chosen for two reasons. First, they are able to deal with both continuous and categorical variables. Second, these algorithms showed good performance in T_{air} interpolations using similar predictor and response variables in other environments (Appelhans *et al.*, 2015). A major

advantage of the machine learning algorithms is that, according to the conceptual designs, they are able to account for different relationships between predictor and response variables under different conditions (e.g., summer/winter). A split into separate models which has been found advantageous in, e.g. Huang *et al.* (2015) is therefore not necessary. James *et al.* (2013) and Kuhn and Johnson (2013a) provide a detailed description of the machine learning algorithms.

The Caret package (Kuhn, 2014a) for R was used as a wrapper package for the GBM (Ridgeway, 2015), RF (Liaw and Wiener, 2002), and Cubist (Kuhn *et al.*, 2014) implementations in R. Models were trained in parallel on 16 cores using the R package "doParallel" (Revolution Analytics and Weston, 2014).

5.2.2.2 Cross-Validation Strategies and Feature Selection to Minimize Overfitting

Overfitting the time series of the training data is a common phenomenon in spatio-temporal prediction models (e.g., Gasch *et al.*, 2015). This means that models can very well predict the time series of the weather stations used for training, but fail in the prediction of "unknown" locations. The term overfitting usually refers to a poor fit of the testing data due to inappropriate model parameters Kuhn and Johnson (2013a). However, though it is rarely approached in literature, we hypothesize that overfitting can also be a result of inappropriate predictor variables.

Overfitting due to inappropriate predictor variables becomes obvious in the difference of the model performance estimated by a random test subset compared to the performance estimated by a Leave-One-Station-Out Cross-Validation (LOSOCV). In order to train a model which is able to successfully predict beyond the location of the training weather stations, a selection of robust variables is required.

Wrapper feature selection methods, that evaluate multiple models, are an intuitive and effective solution to reduce the number of variables to the most important ones Guyon and Elisseeff (2003); Kuhn and Johnson (2013a). However, the most commonly used method for feature selection, recursive feature elimination, relies on variable importance scores which are calculated using the training subset solely (Kuhn and Johnson, 2013a). Thus, recursive feature selection does not account for variable induced overfitting since the subsequent models are based on the ranked variables from the training dataset. If a variable leads to considerable overfitting, it has a high importance in the models. Therefore, this variable will be ranked as an important variable in the recursive feature selection process and is not removed in this process, regardless of a resulting high LOSOCV error.

Therefore, a forward feature selection Guyon and Elisseeff (2003) in conjunction with LOSOCV was applied to remove variables that lead to spatial over-

fitting. We first trained models using all possible 2-variable combinations of predictor variables. The best model of these initial models was kept. The number of predictor variables was then iteratively increased. The improvement of the model was tested for each additional predictor using LOSOCV. We stopped increasing the number of variables when none of the remaining variables decreased the LOSOCV Root Mean Square Error (RMSE) within one standard deviation of the current best model.

Since the process requires considerable computation time, feature selection was only performed using the fastest algorithm, Cubist, and it was assumed that the importance of the variables would be similar for all three algorithms.

To estimate overfitting due to inappropriate predictor variables we compared the performance of the full model (all predictor variables) with the performance of the model that based on the selected variables. We estimated the performance using the LOSOCV predictions. Further, to assess the ability of the model to predict on random test subsets of the data from weather stations used for model training, we predicted on the held out 60% of the overall data set. Overfitting was estimated by comparing the random test subset performance with the LOSOCV performance. RMSE and coefficient of determination (R^2) were used as evaluation scores.

5.2.2.3 Final Model Training, Evaluation and Prediction

The predictor variables retained in the feature selection process were used for training of the final RF, GBM and Cubist models. During model training, the optimal hyperparameters were identified (parameter tuning). Hyperparameters are algorithm specific parameters that cannot be directly estimated from the data but must be specified prior to model training. A majority of the hyperparameters control the model complexity Kuhn and Johnson (2013a). Therefore they must be carefully chosen to avoid overfitting due to highly complex model structures Kuhn and Johnson (2013a). In contrast, a very low complexity might not lead to an optimal fit of the data. To identify the optimal values, models were repeatedly trained using different values for the hyperparameters and the performance was estimated using LOSOCV. While tuning was kept to a minimum (3 different values per parameter) during the time consuming feature selection, the final models were extensively tuned (Table 5.1). See (Kuhn and Johnson, 2013a; James *et al.*, 2013) for a description of the hyperparameters. The optimal values leading to the lowest RMSE based on LOSOCV were used in the final models.

Models were evaluated according to their LOSOCV RMSE. Further, the final models were applied on the overall dataset to assess differences in the performance depending on season and location of the weather stations. Differences were as-

sessed using t-tests. The T_{air} predictions were further aggregated to daily, weekly and monthly estimates by simple averaging of instantaneous predictions.

Table 5.1: *Tested values for the hyperparameters of the different prediction models and the optimal parameters revealed during parameter tuning. N is the number of predictor variables.*

Algorithm	Hyperparameter	Tested Values	Opt. Value
Random Forest	mtry	2 to N with increment 1	2
Cubist	committees	1 to 50 with increment 5	31
	neighbors	0 to 9 with increment 1	0
GBM	number of trees	25 to 500 with increment 25	75
	max depth of interactions	1 to N with increment 1	1
	shrinkage	0.01, 0.1	0.01
	min observations in terminal nodes	10	10

5.3 Results

5.3.1 Selected Features

The variable importance scores of the full models that used all predictor variables revealed the importance of LST to predict T_{air} (Figure 5.2a). Besides LST, terrain-related variables were important, followed by month, season and time of the day. The sensor (Terra or Aqua), as well as the location of the stations on either ice or no-ice had no relevance for the model outcome.

Using the forward feature selection method explained in Section 5.2.2.2, particularly the terrain related variables were identified as leading to overfitting. During forward feature selection, the number of predictor variables was reduced to only LST and month. However, LST was by far the most important predictor in the model (Figure 5.2b). Within one standard deviation of this two-variable model, no further variable could improve the performance.

Figure 5.3 visualizes the problem of spatial overfitting by showing the agreement between measured and predicted T_{air} using two different validation strategies and two different models. Figure 5.3a,b shows the agreement of the full model that uses all predictor variables. Figure 5.3c,d shows the agreement of the model that uses only LST and month as predictor variables. The models were validated using the 60% random subset (Figure 5.3a,c) or using LOSOCV (Figure 5.3b,d). When the full model was applied to the random test subset, it showed a very good fit to the measured T_{air} (RMSE = 6.00 °C, R^2 = 0.78). However, when the model was validated by LOSOCV, the error increased (RMSE = 13.00 °C, R^2 = 0.65) which suggests overfitting. Regarding the 2-variable model, the

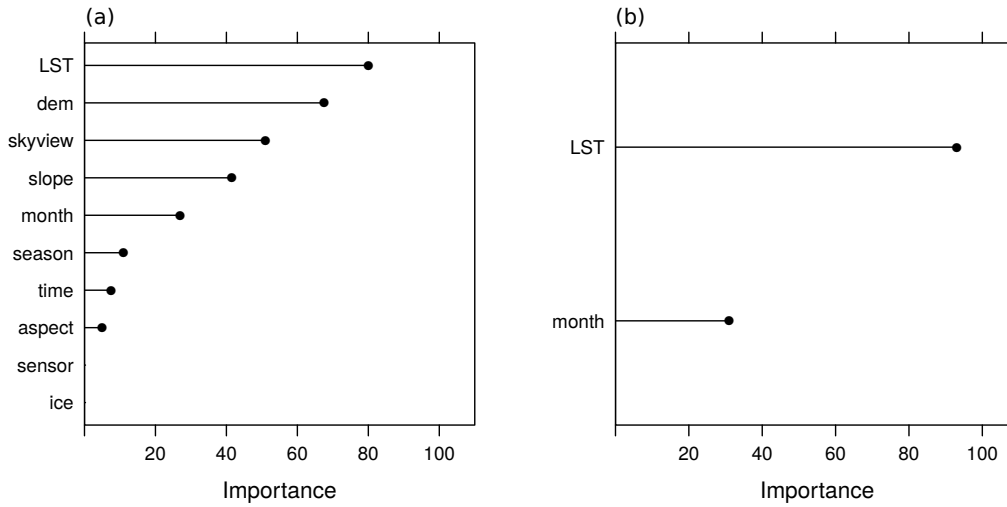


Figure 5.2: *Relative variable importance revealed by Cubist for (a) the full model and (b) the model that used variables selected during forward feature selection only. Variable importance was estimated as the percentage of times each variable was used in the model Kuhn (2014a).*

comparably good results of the full model validated by the random subset could not be kept because the variables that led to overfitting were missing (RMSE = 9.00 °C, $R^2 = 0.71$). When the 2-variable model was validated by LOSOCV, the differences to the random subset validation were less striking (RMSE = 10.84 °C, $R^2 = 0.69$). However, compared to the full model, the RMSE could be decreased by 2.16 °C. This increase in the LOSOCV performance of the forward feature selection based model, highlights the importance of the feature selection to avoid spatial overfitting caused by inappropriate predictor variables.

5.3.2 Model Comparison and Evaluation

The linear model has the form $T_{air} = 0.66 \times \text{LST} - 3.99$. The model was able to predict T_{air} with a LOSOCV R^2 of 0.64 and a RMSE of 11.02 °C. GBM was identified as the best performing algorithm (Figure 5.4) with a R^2 of 0.71 and a RMSE of 10.51 °C. The tuned and trained GBM model was therefore chosen as the final model to create the T_{air} product. Cubist performed slightly worse than GBM (RMSE = 10.85 °C, $R^2 = 0.69$) and the differences to the linear model were small. RF showed the lowest performance among the tested algorithms (RMSE = 11.95 °C, $R^2 = 0.56$).

In the following, we focus on the GBM model, since this model was applied in the creation of the final T_{air} product. In order to further assess characteristics of the model predictions, the model was applied to the full dataset.

The interquartile ranges of the measured and predicted data were similar (Figure 5.5). However, the model was not able to predict very high or very low

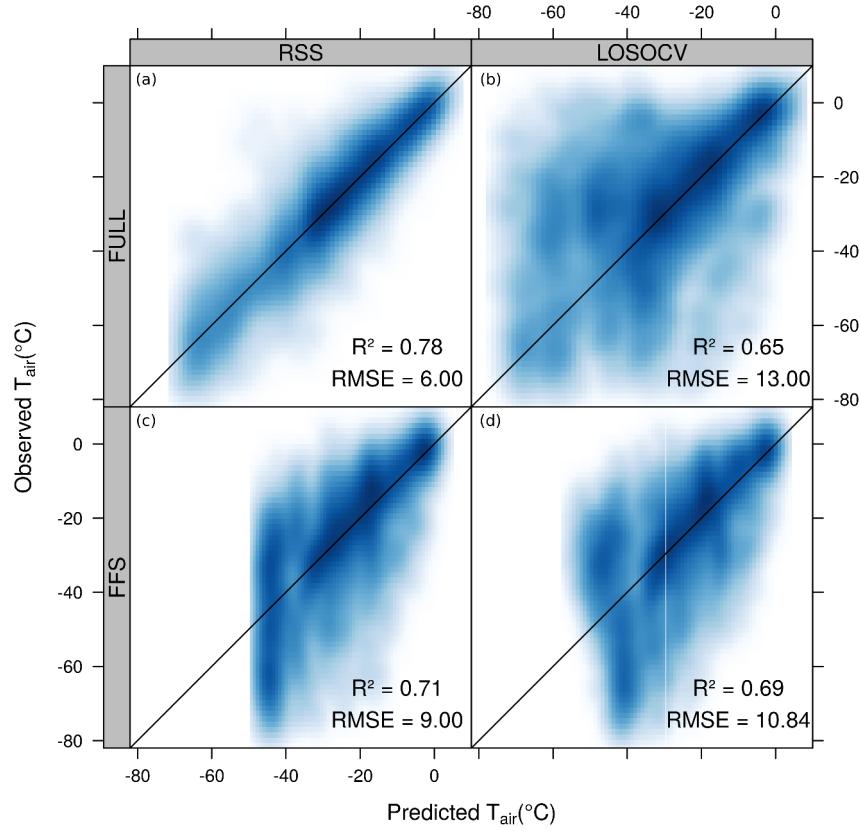


Figure 5.3: Agreement between measured T_{air} and predicted T_{air} using Cubist as modeling tool. Models were first trained using the full set of predictor variables (FULL) and second using only LST and month as predictors that were revealed during forward feature selection (FFS). Two different validation strategies were used: Comparison to a 60% random subset of the total dataset (RSS), as well as Leave-One-Station-Out Cross-Validation (LOSOCV). The problem of overfitting becomes obvious by comparing the performance of the full model validated by RSS (a) with the performance of the full model validated by LOSOCV (b). Using the 2-variable model, the performance validated by RSS decreased (c) compared to the full model. Only slight differences between RSS and LOSOCV (c,d) indicate that there is no considerable overfitting in the 2-variable model. The most important point is that the 2-variable model could improve the LOSOCV performance (d) compared to the full model. Data are represented as smoothed scatter plots where the intensity of the color represent the data point density.

values. Since low T_{air} values (<40 °C) were still frequently measured by the weather stations, the main weakness of the model was the inability to predict low temperatures.

Figure 5.6 shows a comparison between the measured and predicted time series of example weather stations (Figure 5.6a–c), as well as of the average time series of all 32 weather stations (Figure 5.6d). The seasonal T_{air} patterns were generally well captured both by the LST and by the predicted T_{air} time series. Compared to the linear model, the GBM model could differentiate between summer and winter

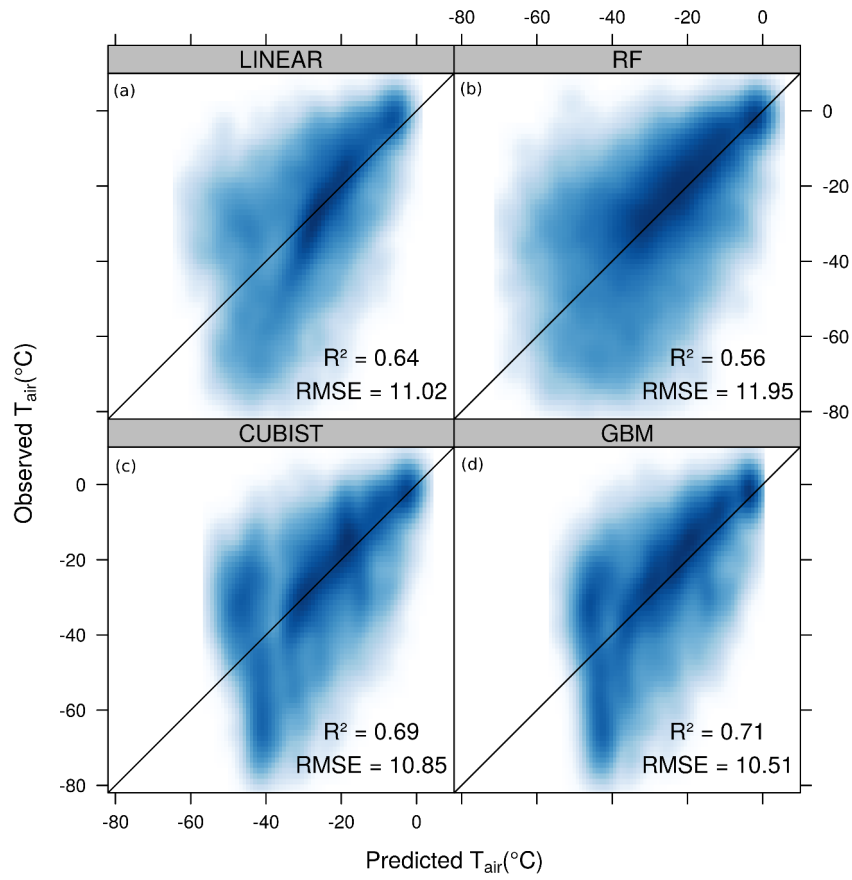


Figure 5.4: Correlation between measured and predicted T_{air} based on LOSOCV of (a) the linear model, (b) RF, (c) Cubist and (d) GBM. The intensity of the color represent the data point density.

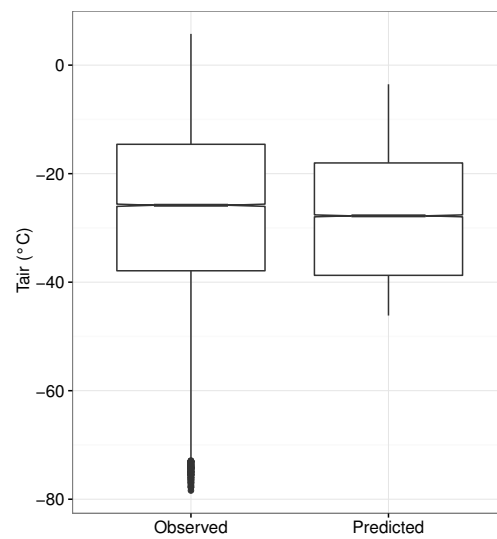


Figure 5.5: Distribution of measured T_{air} compared to predicted T_{air} .

months and was therefore closer to the measured time series of the McMurdo Dry Valleys (Figure 5.6a,b), as well as of the average time series. During winter, no clear advantage of GBM over the linear model could be observed for the McMurdo Dry Valleys (Figure 5.6a,b). Both models fail in the prediction of the time series of "Harry" which is located on ice and has considerably lower measured temperature values (Figure 5.6c). Regarding the average time series, GBM outperformed the linear model during both, summer and winter.

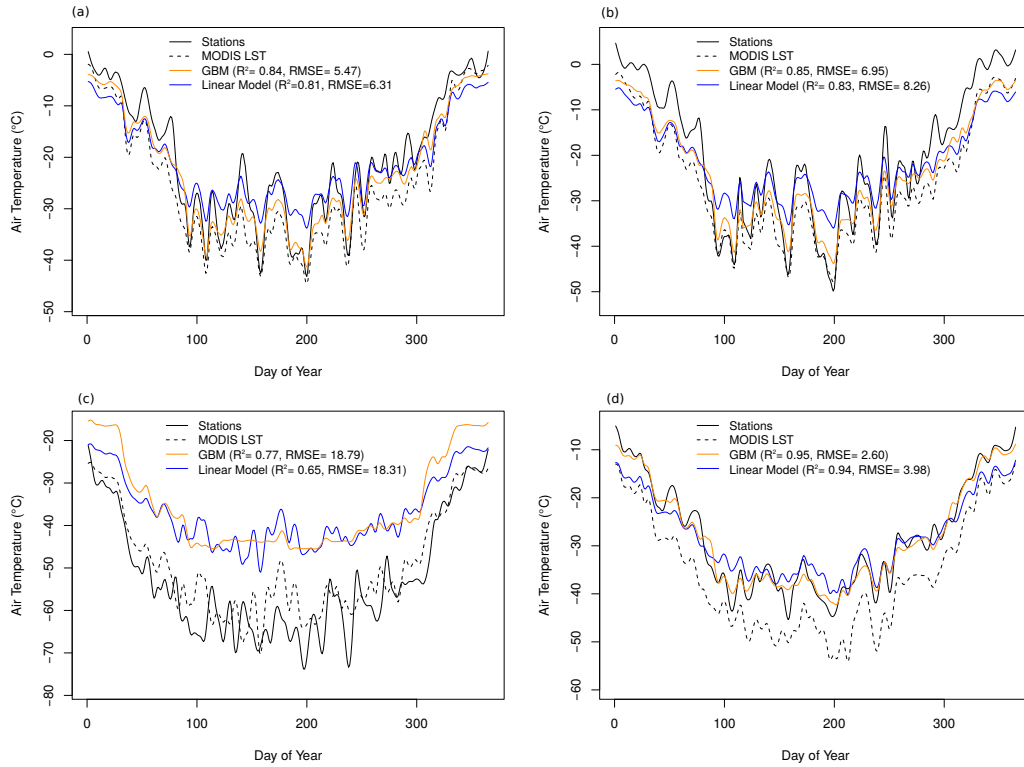


Figure 5.6: Time series of the three example weather stations (a) "Brownworth", (b) "Bull pass" and (c) "Harry", as well as (d) the average time series of all 32 weather stations for the year 2013. "Brownworth" and "Bull pass" are weather stations located in the McMurdo Dry Valleys. "Harry" is located on ice. The lines represent the measured T_{air} (solid black), MODIS LST (dotted black), GBM model predictions (orange), as well as linear model predictions (blue).

On average, the RMSE between measured and predicted values was significantly higher in the winter months (e.g., Jul, Aug) than in the summer months (e.g., Jan, Dec) (Figure 5.7). This holds especially true for the weather stations that are located on ice-free areas such as the McMurdo Dry Valleys (Figure 5.7b). In general, the errors were significantly lower for the weather stations in ice-free areas compared to the stations located on ice ($p < 0.01$) (Figures 5.7 and 5.8). Regarding the stations located on ice, those in the west and south had the lowest RMSE. Apart from this observation, no clear spatial patterns could be observed for these stations (Figure 5.8).

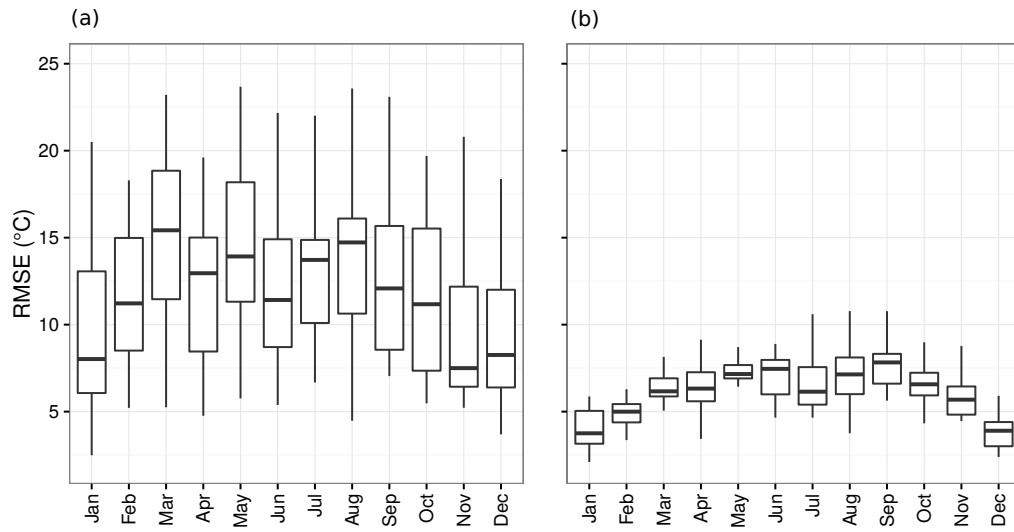


Figure 5.7: Monthly RMSE of T_{air} predictions separately for (a) weather stations located on ice and (b) weather stations located on ice-free areas.

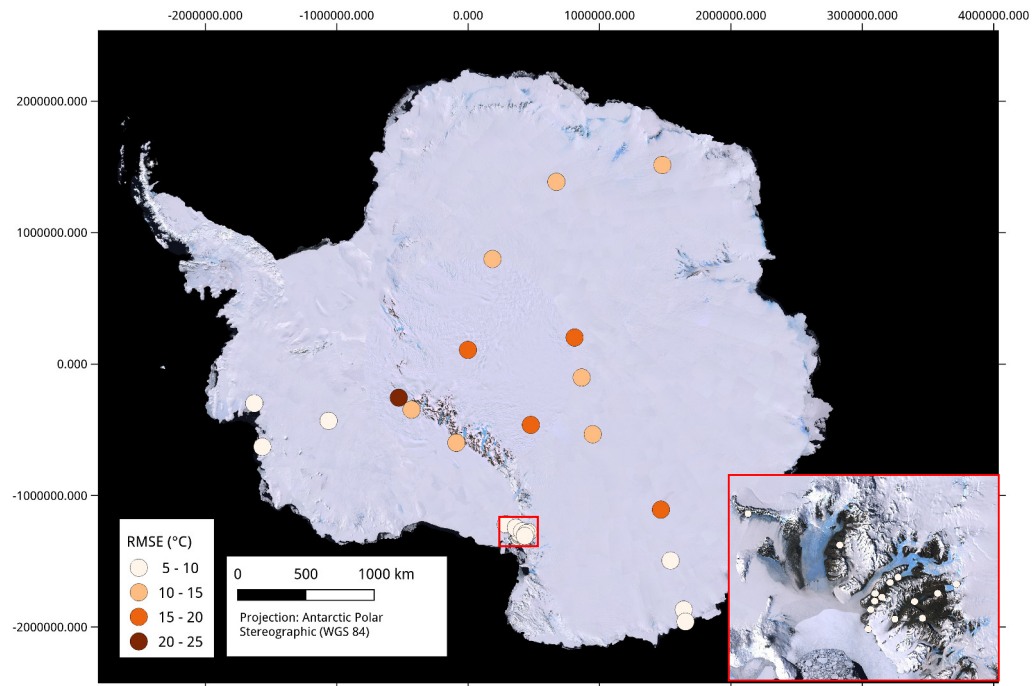


Figure 5.8: Spatial distribution of RMSE of predicted T_{air} for the 32 weather stations. The McMurdo Dry Valleys are shown in detail. The weather stations are overlaid on a Landsat composite image (U.S Geological Survey, 2007).

When the product was aggregated to daily, weekly or monthly data the agreement between measured and predicted T_{air} increased with the aggregation level ($p < 0.01$, Fig. 5.9). Figure 5.10 shows the final product aggregated to monthly data.

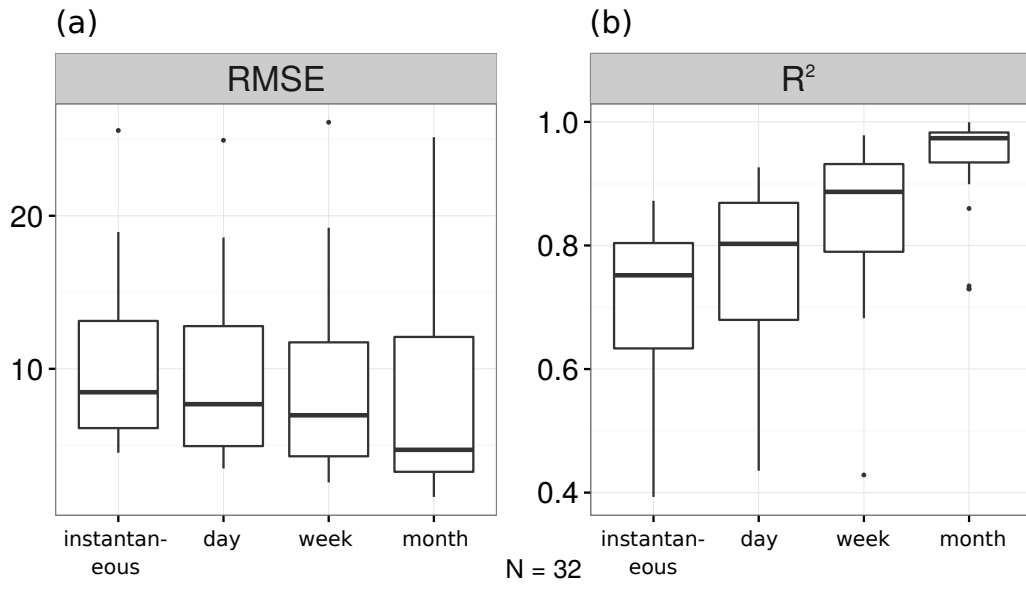


Figure 5.9: Performance of the T_{air} product on different aggregation levels.

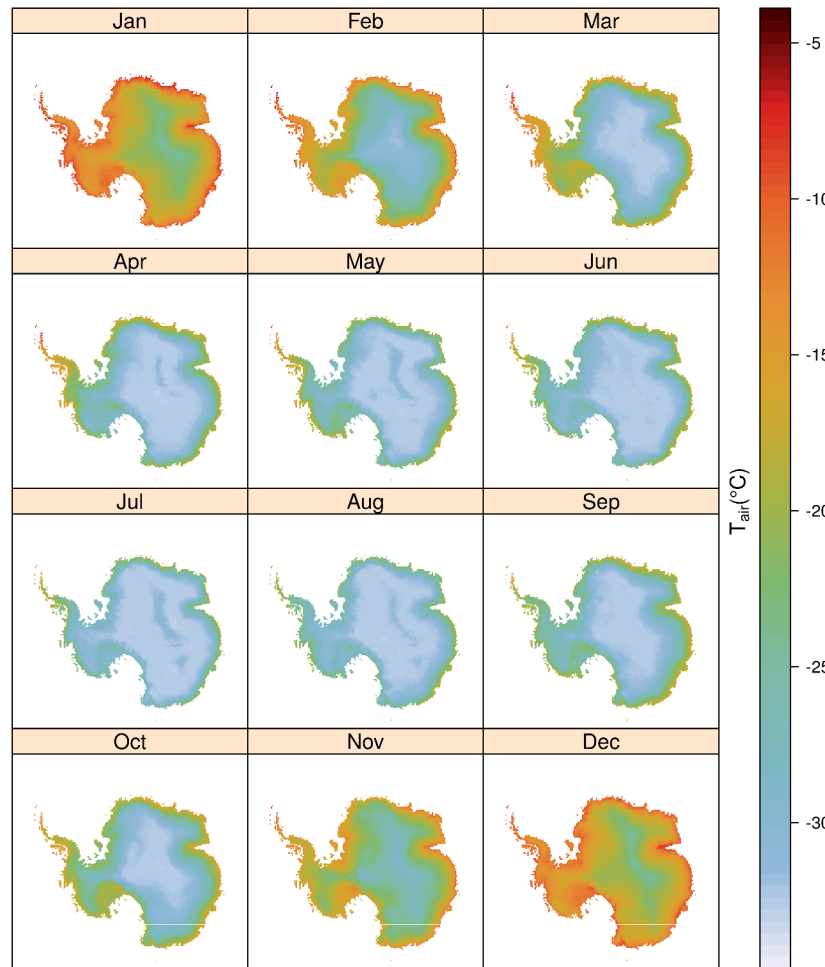


Figure 5.10: Monthly aggregates of T_{air} as predicted by the GBM model.

5.4 Discussion

GBM using LST and month as predictors was the best performing model. Nevertheless, the simple linear estimation of T_{air} from LST is competitive in Antarctica where no considerable land cover differences occur. However, the machine learning algorithms including the month as additional predictor were better able to approximate the time series of the weather stations.

The weakness of the model was the prediction of very low temperatures (approx. below -35°C), where the accordance between measured and predicted T_{air} decreased. A direct comparison of the performance of the Antarctica T_{air} product to other studies is difficult, not only due to differing environments and temporal scales, but most particularly due to different validation methods being used. Benali *et al.* (2012); Hengl *et al.* (2011); Shi *et al.* (2016) found considerably higher agreement between measured and predicted T_{air} by including auxiliary variables in their models (RMSE usually below 2.5°C). However, their model training and validation strategies do not rely on LOSOCV. Since LOSOCV is considered to be a stricter validation strategy (Gasch *et al.*, 2015), it is not surprising that the agreement was better in these studies. Keeping in mind that the model performance in our study was considerably higher without LOSOCV compared to the LOSOCV performance (Figure 5.3a,b), it is likely that using a similar validation approach would lead to less divergent results.

Forward feature selection in conjunction with LOSOCV allowed removing variables that led to overfitting. Particularly the terrain related variables caused this problem. One characteristic of these variables is that they change in space but not in time which means that each weather station has a unique combination of these variables. We assume that these "static variables" are prone to overfitting since they are overrepresented in the predictor dataset. The weather station dependent combination of unique properties is quasi-comparable to an ID of the stations which is then used as predictor variable. Using an ID as predictor, the model would be able to fit general characteristics of the individual time series which are, however, not valid for unknown locations. Therefore, these variables need to be checked using feature selection in conjunction with LOSOCV and removed if they are misleading.

Though the LOSOCV performance could be improved using selected predictor variables only, the new T_{air} product is afflicted with considerable errors that need to be explained. A first explanation comes from the characteristic of machine learning algorithms which are not able to predict extreme values (i.e., very low and very high temperatures) Kuhn and Johnson (2013a). However, since the model also showed high errors in the T_{air} range that was well represented in the training data, this probably only slightly contributed to the overall error.

It could be shown that the weather stations located in ice-free areas could be predicted with much higher accuracy compared to the weather stations located on ice which could not be solved by including the information of ice as predictor variable. Though the MODIS LST product is cloud masked, we realized that cirrus clouds could not reliably be removed from the data. Due to the "white on white" and "cold on cold" problem (Bromwich *et al.*, 2013; Allen Jr. *et al.*, 1990), cloud classification over Antarctica is challenging in winter months and over snow or ice covered areas. This causes T_{air} predictions in winter months and over snow and ice covered areas more prone to errors than in summer months and/or over ice-free areas. Also Janatian *et al.* (2017) reported a significant decrease of model performance to predict T_{air} in low temperatures which might also be a result of inaccurate cloud masking in cold environments.

Another issue that likely affects the model performance is that, due to the remoteness of Antarctica, the maintenance of weather stations is challenging, and the number of stations available for ground truthing is very limited. Due to the difficulty of maintenance, the data are likely afflicted with higher errors than those of less remote areas. Errors in the data used as ground truth can have a high impact on the model outcome. An extensive quality check of the weather station data would be important in future extensions of this study.

We found no systematic patterns of accuracy for the stations located on ice. This suggests that the causes for the low performance are due to local influences rather than due to systematic errors in the LST product. Very local microclimatic influences that cannot be captured by the 1km resolution are a potential explanation but also the suggested errors in the station data. Therefore, a station-specific assessment of error sources is a future task in view to an improvement of the results.

Despite the errors, the T_{air} product is of high value for scientific studies in Antarctica.

The advantage is the high temporal (sub-daily to daily), as well as spatial resolution which could only be achieved using remote sensing data. The estimates provide instantaneous T_{air} information at the time of overpass of the Terra and Aqua satellites. These instantaneous estimates are useful to feed models that require high temporal resolution T_{air} estimates. However, a variety of studies might be more interested in temporal aggregates of the product, such as on weekly or monthly scales. In this context it is of note that the simple averaging to daily composites, as performed in this study, could be improved by considering the diurnal T_{air} cycle (Jin and Dickinson, 1999).

The product allows to monitor the spatio-temporal dynamics of T_{air} , not only on a continental scale, but also in a regional scale for example for the McMurdo Dry Valleys. The T_{air} product can be considered as a baseline dataset to

understand the regional and local climate variability of Antarctica. This especially applies to research areas requiring a spatially coherent gridded dataset for regional climate model evaluation in terms of linking local meteorological processes, such as topographically induced warming and cooling events, to non-local atmospheric circulation patterns (such as low pressure systems). However, it is of note that the product is currently not suitable to analyse patterns of small temperature changes. As an example, the estimated increase of T_{air} by 2.4 ± 1.2 °C over the West Antarctic Ice Sheet since the 1950s Bromwich *et al.* (2013) will most likely not be captured. Against the background of the relevance of climate change, an improvement of the product will be required. However, a variety of ecological studies focus on the ice-free areas of Antarctica and on the summer months where organisms are active (Convey *et al.*, 2014). In this context, the product provides T_{air} estimates with acceptable errors, especially when temporal aggregates (weekly, monthly) are considered.

The model was trained using data from 2013 only. To extend the model over the entire MODIS lifespan (MODIS LST is available since 2000), it will be necessary to include a subset of data from further years for model training to ensure that a wider range of inter-annual environmental conditions is included. Since the month was revealed as the second important predictor variable, it is important to include data from further years into training to avoid an overfitting to this specific year. At this point, the advantage of the linear model is of note which has shown a comparable performance. Since this model relied on LST solely, and we presume that the relation between LST and T_{air} is not strongly affected between interannual changes (except for sensor degradation), it can directly be applied beyond the training year 2013.

It is of note that the relationship between LST and T_{air} is influenced by various other parameters that could not be employed in this study. Mean wind speed and direction could be responsible for horizontal heat advection, while near-surface wind turbulence allows for surface-atmosphere energy exchange through the sensible and latent heat flux. In further studies, it might be worth to combine the LST data with regional climate model results that estimate parameters such as the surface energy balance components (net surface radiation, sensible, latent and ground heat flux).

5.5 Conclusions

A methodology was presented to predict T_{air} from MODIS LST and auxiliary data using machine learning algorithms. LST and month of the year were the most suitable and robust predictors for T_{air} . Among the tested models, GBM was the most promising algorithm. However, the differences to the commonly

used simple linear approach were rather small. The main weakness of the model was failing to predict extremely low temperatures (e.g., below -35°C). The T_{air} estimates are in 1km spatial and daily temporal resolution for the entire continent for the year 2013. The product is available from the authors on request. Future research needs to focus on minimizing errors followed by an extension of the product to the overall lifespan of the MODIS sensor.

Acknowledgments This research was supported by the Ross Sea Region Terrestrial Data Analysis research program, funded by the Ministry of Business and Innovation, New Zealand, with contract number CO9X1413. The authors would like to thank the anonymous reviewers for their careful attention to this work, and the New Zealand eScience Infrastructure (NeSI) for the computational resource support.

Chapter 6

From local spectral measurements to maps of vegetation cover and biomass on the Qinghai-Tibet-Plateau: Do we need hyperspectral information?

Hanna Meyer (1), Lukas W. Lehnert (1), Yun Wang (2), Christoph
Reudenbach (1), Thomas Nauss (1), Jörg Bendix (1)

(1) Faculty of Geography, Philipps-University Marburg, Deutschhausstr. 10, 35037 Marburg,
Germany

(2) Senckenberg Museum of Natural History, Am Museum 1, 02826 Görlitz, Germany

Published in

International Journal of Applied Earth Observation
and Geoinformation, 2017, 55, 21–31

Received 3 March 2016

Revised 8 August 2016

Accepted 1 October 2016

Available online 24 October 2016

Published March 2017

6 From local spectral measurements to maps of vegetation cover and biomass on the Qinghai-Tibet Plateau: Do we need hyperspectral information?

Hanna Meyer, Lukas W. Lehnert, Yun Wang, Christoph Reudenbach, Thomas Nauß, Jörg Bendix

Abstract

Though the relevance of pasture degradation on the Qinghai-Tibet Plateau (QTP) is widely postulated, its extent is still unknown. Due to the enormous spatial extent, remote sensing provides the only possibility to investigate pasture degradation via frequently used proxies such as vegetation cover and aboveground biomass (AGB). However, unified remote sensing approaches are still lacking. This study tests the applicability of hyper- and multispectral *in situ* measurements to map vegetation cover and AGB on regional scales. Using machine learning techniques, it is tested whether the full hyperspectral information is needed or if multispectral information is sufficient to accurately estimate pasture degradation proxies. To regionalize pasture degradation proxies, the transferability of the locally derived ML-models to high resolution multispectral satellite data is assessed. 1183 Hyperspectral measurements and vegetation records were performed at 18 locations on the QTP. Random Forests models with recursive feature selection were trained to estimate vegetation cover and AGB using narrow-band indices (NBI) as predictors. Separate models were calculated using NBI from hyperspectral data as well as from the same data resampled to WorldView-2, QuickBird and RapidEye channels. The hyperspectral results were compared to the multispectral results. Finally, the models were applied to satellite data to map vegetation cover and AGB on a regional scale. Vegetation cover was accurately predicted by Random Forest if hyperspectral measurements were used (cross validated $R^2 = 0.89$). In contrast, errors in AGB estimations were considerably higher (cross validated $R^2 = 0.32$). Only small differences in accuracy were observed between the models based on hyperspectral compared to multispectral data. The application of the models to satellite images generally resulted in an increase of the estimation error. Though this reflects the challenge of applying *in situ* measurements to satellite data, the results still show a high potential to map pasture degradation proxies on the QTP. Thus, this study presents robust methodology to remotely detect and monitor pasture degradation at high spatial resolutions.

Keywords Pasture degradation; Qinghai Tibet Plateau; Hyperspectral measurements; Regionalization; Random Forests; Biomass; Vegetation cover

6.1 Introduction

Livestock grazing is the dominant land-use of the grasslands on the Qinghai-Tibet Plateau (QTP). The pastures are the economical basis for the Tibetan people, providing forage for approximately 13 million yaks and 41.5 million sheep (Long *et al.*, 1999) and have been formed by thousands of years of pastoralism (Miehe *et al.*, 2009). However, in the last decades the Tibetans are faced with increasing grazing pressure caused by increasing population numbers (Harris, 2010). Pasture degradation due to over-grazing is presumably the consequence which is strengthened by climate change (Lehnert *et al.*, 2016) and improper grazing management (Cao *et al.*, 2013). The degradation of the pastures is of significant economic importance for the population. In addition, it affects even larger scale patterns such as the discharge flow rates and sediment redistribution in major river catchments (Asner *et al.*, 2004) or an alteration of radiation feedback (Gong Li *et al.*, 2000).

Values for the extent of pasture degradation on the QTP vary heavily, because studies are conducted in a subjective way, are poorly documented, or cover only small spatial extents (Harris, 2010). To overcome these spatial deficits, remote sensing based approaches are needed. Therefore, satellite images with a high spatial resolution taken with sensors like WorldView-2 (WV), RapidEye (RE) or Quickbird (QB) must be analyzed. Due to the large spatial extent and the difficult access to the area, semi-automatic approaches requiring very little field data would be advantageous.

Proxies for pasture degradation on larger scales were successfully derived from multi- (Wessels *et al.*, 2008; Zha *et al.*, 2003) as well as hyperspectral (Beeri *et al.*, 2007; Huang *et al.*, 2004; Lehnert *et al.*, 2014) remote sensing data. While vegetation cover is the most frequently used proxy for pasture degradation in multispectral approaches (Gao *et al.*, 2010; Liu *et al.*, 2004; Lehnert *et al.*, 2015b), hyperspectral data were used to quantify further proxies for pasture degradation including biomass (Itano and Tomimatsu, 2011), species diversity (Fava *et al.*, 2010) and chemical foliage composition (Kokaly and Clark, 1999; Lehnert *et al.*, 2013, 2014). The most commonly considered hyperspectral indices to estimate these proxies encompassed shape and size of absorption features (Mutanga and Skidmore, 2004), red edge parameters (Mutanga and Skidmore, 2007), vegetation indices as well as narrow band indices (Thenkabail *et al.*, 2002). Though some studies compared indicators derived by spectrometer measurements with satellite-derived indices (Numata *et al.*, 2008) there has been little research on applying revealed relationships between the spectrometer-derived hyperspectral indices and proxies for pasture degradation on larger scale multispectral satellite images (Liu *et al.*, 2004; Psomas, 2008).

This study is aimed at assessing the suitability of hyper- and multispectral remote sensing data to regionalize proxies for pasture degradation on the QTP. Therefore,

1. it should be tested if a high spectral resolution of hyperspectral data compared to multispectral approaches considerably improves the estimation accuracy of machine learning models to derive vegetation cover and above-ground biomass (AGB) as proxies for pasture degradation, and
2. the usability of spectrometer measurements should be assessed to derive models applicable to regionalize pasture degradation proxies based on high spatial resolution multispectral satellite images.

6.2 Data and methods

A valid and comprehensive dataset of field observations to ensure a thorough training and validation of the derived parameter estimations is a prerequisite to perform a regionalization of pasture degradation proxies on the QTP. Therefore, this section is divided into a short description of the study area followed by the acquisition of field data and the description of the satellite data. The main part describes the derivation and regionalization of pasture degradation parameters using hyper- and multispectral data.

6.2.1 Study area

Alpine meadows and alpine steppes are the dominating vegetation types on the QTP (Hou, 2001) (Fig 6.1). Alpine meadows can be found at altitudes between 3500 m and 4500 m a.s.l. where annual precipitation exceeds 400 mm (Sheehy *et al.*, 2006). They usually feature a closed vegetation cover which is unique among the vegetation of the QTP (Miehe, 2004). Alpine meadows are composed of plants of the genus *Carex* (Miehe *et al.*, 2008b) (partly former *Kobresia*, Global Carex Group, 2015), where the dominant species *Carex parvula* typically forms a thick turf-layer protecting the soil surface against erosion (Miehe *et al.*, 2008b). In degraded areas, the turf is less intact (Miehe *et al.*, 2011b). Alpine steppes are dominated by species of the genus *Stipa*. The vegetation cover is typically less than that of alpine meadows and does not have a sod layer. In addition to grasses, cushion plants characterise this vegetation type (Miehe, 2004; Sheehy *et al.*, 2006). For a detailed description of the vegetation types refer to Miehe *et al.* (2008a,b); Miehe (2004); Sheehy *et al.* (2006).

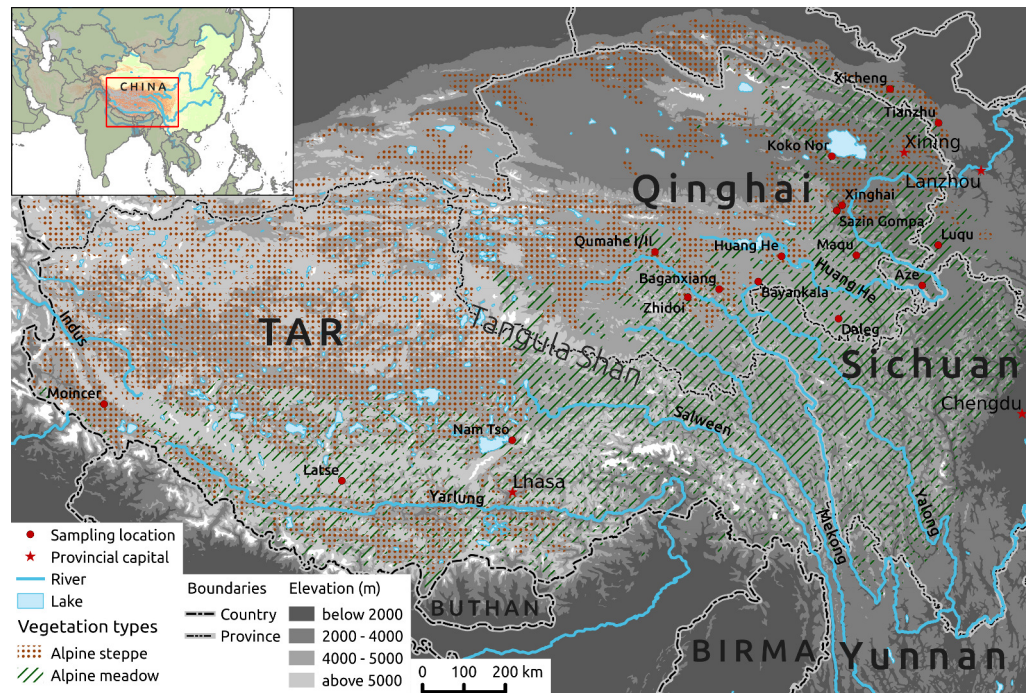


Figure 6.1: Map of the study area including vegetation types considered in this study and the position of the 18 sampling locations. The distributions of vegetation types are from Hou (2001) and Miede et al. (2008b).

6.2.2 Field work

To record a comprehensive dataset regarding the spectral and degradation proxies, 18 sampling locations on the QTP were investigated in 2011 and 2012 between mid June and early September, when the growth of vegetation is highest on the plateau (Fig. 6.1, Tab. 6.1). The sites were chosen with the aim to cover different areas of alpine steppes and alpine meadows and were restricted to accessibility rights.

Hyperspectral measurements and RGB photography to obtain vegetation cover were taken following grazing gradients starting at a small villages or camp sites and moving away from them. The length of the transects was variable depending on the distance to neighbouring settlements or was finished when fences led to a rapid change in gradients. The direction of the transects was chosen such that the aspect and slope remained fairly constant throughout the transect. Along each transect, measurements were performed in predefined distances. The distance between sites was 15 m in the first third of the transect. Moving away from the settlement, we increased the increments to 30 m in the second and 60 m in third part of the transect to account for the non-linearity of cattle density with increasing distance from the settlements (Sternberg, 2012) which affects vegetation cover and biomass. At each location, a representative site near the settlement and a site at the end of the transect was determined for biomass measurement.

Table 6.1: Overview of the 18 locations which were sampled during field work in 2011 and 2012.

Site	Lat.	Lon.	Year	Vegetation	Number of	
				Type	Spectral measurements	AGB samples
Aze	101.89	33.7	2011, 12	meadow	147	1
Baganxiang	96.56	33.96	2012	meadow	188	6
Bayankala	97.61	34.08	2011	meadow	4	0
Daleg	99.64	33.15	2012	meadow	77	2
Huang He	98.26	34.6	2011	steppe	16	0
Koko Nor	99.81	36.7	2011	meadow	5	0
Latse	86.85	29.38	2012	meadow	17	0
Luqu	102.42	34.54	2011	meadow	3	0
Maqu	100.23	34.49	2011	meadow	6	0
Moincer	80.73	31.15	2012	steppe	110	5
Nam Tso	91.13	30.79	2012	meadow	191	6
Qumahe I	94.92	34.85	2012	meadow	237	2
Qumahe II	94.99	34.88	2012	steppe	88	1
Sazin Gompa	99.82	35.5	2011	meadow	6	0
Tianzhu	102.79	37.19	2011	steppe	3	0
Xicheng	101.57	38.04	2011	steppe	6	0
Xinghai	99.98	35.61	2011	steppe	8	0
Zhidoi	95.74	33.82	2011, 12	meadow	71	2

6.2.2.1 Spectral measurements

Hyperspectral measurements were taken using a HandySpec Field spectrometer (Tec5 AG, Oberursel, Germany) measuring from 305 to 1705 nm with a spectral resolution of 1 nm. The device has two channels which cross at 1050 nm. Spectrometer measurements were conducted between 10 am and 5 pm using direct and indirect radiation from the sun as light source. Measurements were calibrated with a white panel every time the spectrometer was turned on. In addition, the device was recalibrated if light conditions changed. Reflectance was measured from small subsets of the ecosystem (termed "site" in the following), corresponding to a circle of 20 cm in diameter at every 10 meters on the transect. Sandmeier *et al.* (1998) investigated BRDF effects of grass and found anisotropic factors of approximately 1 for a similar measurement setup. Because the vegetation height is considerably lower at the QTP than in the study of Sandmeier *et al.* (1998), we did not correct our hyperspectral samples for bi-directional effects. In total 1183 spectra were sampled at the 18 locations. The exact geographical position of all sites was recorded using a differential GPS (Topcon HiPer II) in postprocessing mode.

6.2.2.2 Proxies for pasture degradation

Vegetation cover of the field of view of each hyperspectral measurement was assessed by RGB image analysis. Therefore a common digital camera was positioned at a fixed height directly under the spectrometer sensor and one image covering the field of view of the sensor was taken. Vegetation cover was then assessed by performing a simple threshold classification. See Lehnert *et al.* (2015b) for further information on this method. The percentage of green vegetation pixels in the image (i.e. vegetation cover) was used as one of the response variables in the model. The AGB was sampled on 25 of the measured sites. Therefore, the AGB on 0.5 x 0.5 m plots was completely cut directly above the surface. The harvest was air dried and weighted. See Tab. 6.2 for a summary of the vegetation cover and AGB data.

Table 6.2: *Summary statistics of the vegetation cover and AGB samples which were used as response variables in this study.*

Summary Statistic	Vegetation cover (%)	AGB (g/m ²)
Minimum	0.1	14.7
Maximum	100.0	112.9
Mean	55.0	37.4
Standard deviation	28.1	24.2

6.2.3 Satellite data

QB, RE and WV images were ordered in advance of the field work for each of the sampling locations. The data had spatial resolutions of 2 m (QB), 2 m (WV) and 5 m (RE) respectively. QB features four channels, RE has an additional channel in the red edge and the 8 channels of WV nearly cover the whole spectrum from 400 to 1000 nm with a high spectral resolution in the red edge (Tab. 6.3). The final set of satellite images are shown in Tab. 6.4.

6.2.4 Methodology to remotely derive pasture degradation parameters

All processing steps were performed using R Version 3.2 (R Core Team, 2015) if not mentioned otherwise. Hyper- and multispectral functionality was provided by the R package "hsdar" (Lehnert *et al.*, 2015a). The R packages "caret" (Kuhn, 2014a) and "randomForest" (Liaw and Wiener, 2002) were used for training of the machine learning models, and the "raster" package (Hijmans, 2015) was used in the upscaling process. Fig. 6.2 gives an overview about the individual processing steps which are described in the following.

Table 6.3: *Overview of the spectral bands of the three satellite sensors QuickBird, RapidEye and WorldView-2.*

Channel	QuickBird		RapidEye		WorldView-2	
	from	to	from	to	from	to
1	450	520	440	510	401	453
2	520	600	520	590	448	508
3	630	690	630	685	511	581
4	760	900	690	730	589	627
5			760	850	629	689
6					704	744
7					772	890
8					862	954

Table 6.4: *Summary of the satellite images which were available for this study.*

Location	Sensor	Acquisition Date
Aze	WorldView-2	2012-09-02
Huang He	RapidEye	2011-08-28
Moincer	QuickBird	2012-07-23
Namco	QuickBird	2012-09-12
Qumahe	QuickBird	2012-09-05
Tianzhu	WorldView-2	2011-09-12
Zhidoi	WorldView-2	2012-08-13

6.2.4.1 Calculation of predictor feature spaces

The study aims to compare the suitability of hyper- and multispectral remotely sensed data to estimate pasture degradation parameters. Therefore, the hyperspectral data acquired in the field were first resampled to QB, RE and WV channels. Spectral resampling has been performed using the spectral response functions of the respective sensors. The resulting simulated multispectral data are not affected by side effects typical for satellite data such as atmospheric influences or the delay between field and satellite data acquisitions. This is ensured because of the small distance between sensor and object on the ground as well as the instantaneous measurement of vegetation cover (using the instantaneous RGB image) or biomass (cutting at the same day as measurements were taken). The multispectral information used for the models were not taken from the satellite images directly since vegetation cover and AGB were sampled on a scale which was smaller than the resolution of one pixel. This approach allows to directly compare hyper- and multispectral results since the data source is identical. From the hyperspectral and from each multispectral dataset, separate feature spaces

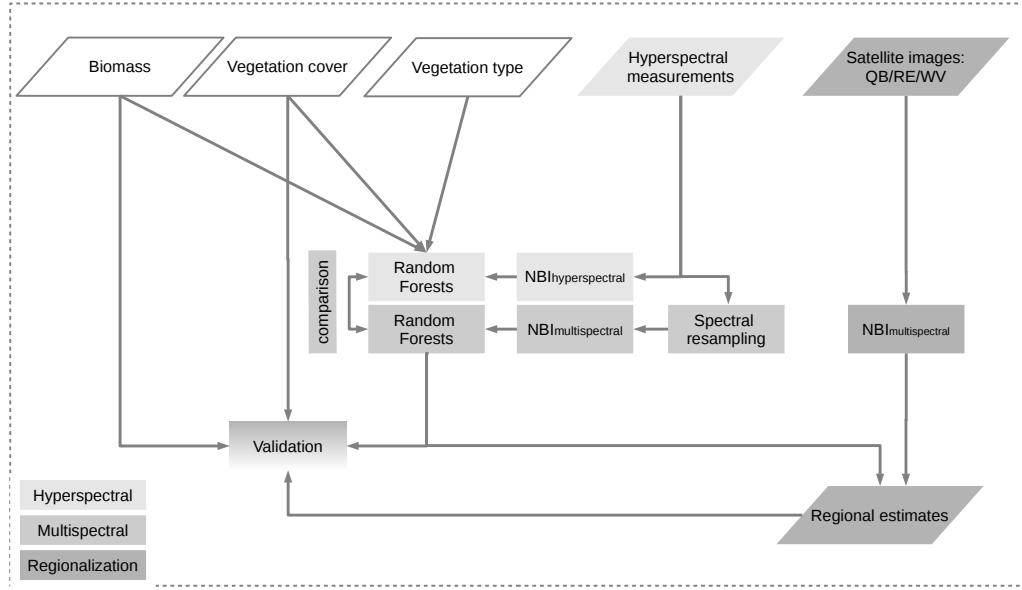


Figure 6.2: *Schema of the processing flow. The white boxes represent datasets taken in the field. The gray colors depict different processing steps to answer the main objectives. Rectangles are processing steps and rhombs are datasets.*

were compiled. For comparability to the multispectral satellite images, only the range from 305 to 1000 nm was considered in the hyperspectral dataset. The feature spaces consisted of normalized band indices similar to the normalized difference vegetation index (NDVI) but including all possible band combinations (Thenkabail *et al.*, 2002; Psomas *et al.*, 2011):

$$NBI_{\lambda_i, \lambda_j} = \frac{R_{\lambda_i} - R_{\lambda_j}}{R_{\lambda_i} + R_{\lambda_j}} \quad (6.1)$$

Here, R is the reflectance at the wavelength λ_i or λ_j . As an additional predictor variable, the vegetation type at the respective sampling location was included as indicated by Hou (2001) and verified or adapted by expert knowledge. For AGB modeling, vegetation cover was used as a second additional predictor variable.

6.2.4.2 Estimation of pasture degradation parameters using Random Forests

Machine learning algorithms such as Random Forests (RF) (Breiman, 2001a) are known as being able to deal with complex interacting as well as highly correlated predictor variables as it is the case regarding spectral bands of optical sensors. A valuable overview as well as practical guide for the usage of RF can be found in Kuhn and Johnson (2013a) and James *et al.* (2013). Though other machine learning algorithms as for example neural networks or support vector machines might perform equally well, RF is an intuitive, computationally efficient and easy to use method which does not require any preprocessing of the variables.

In this study RF was used in conjunction with recursive feature elimination. Recursive feature elimination fits the model first with all predictor variables, i.e. with all variables in the predictor feature spaces. It then calculates variable importance according to Liaw and Wiener (2002) and removes the least important variables. In the next step the model is re-calculated with the updated set of variables. This step is repeated for different numbers of variables. The best number and combination of variables is then determined by comparing the performance of the individual models.

As settings for each RF model, a number of 500 trees was used. The number of predictor variables randomly selected at each split (`mtry`) was tuned between 2 and the number of predictor variables in the respective feature space as suggested by Kuhn and Johnson (2013a). Since especially the AGB data set is rather small, an independent test data set would not be truly representative for the overall data set. Therefore, instead of splitting the data into training and test data, a leave-one-location-out cross validation for vegetation cover modeling and a 50 times repeated 3 fold cross validation for AGB modeling was performed to assess the error of estimation. Leave-one-location-out means that models are repeatedly trained by leaving the data from one of the 18 locations out. Therefore, the size of the validation data varied between 3 and 237 samples. The vegetation cover of the remaining station was then predicted and compared to the measured data to estimate the models ability to predict vegetation cover at unknown locations. The error assessment of the AGB modeling could not be performed using a leave-one-location-out cross validation due to the low sample size. A repeated three folds cross validation to account for the small size of the data set (25 samples) was calculated instead. Thus, the dataset was randomly partitioned into 3 equally sized folds. Consequently, the size of the validation data for each model run was between 8 and 9 samples. Models were then iteratively trained by leaving one of the folds out and predicting on the remaining fold. The procedure was repeated 50 times.

Separate models were trained for the hyper- and each multispectral data of QB, RE and WV. As predictors, all variables in the respective feature space that were identified as important variables during feature selection were used. The cross validated R^2 as well as the root-mean-square error (RMSE) values were compared among the hyper- and the multispectral models.

6.2.4.3 Regionalization

The satellite images were preprocessed prior to the usage for regionalization of degradation parameters. Preprocessing included radiometric, geometric and atmospheric correction. The atmospheric correction was performed using the 6S code (Kotchenova *et al.*, 2006) which was extended as described in Curatola

Fernández *et al.* (2015). See Lehnert *et al.* (2015b) for a detailed description of the parametrization of the 6S-correction for the QTP as well as all other steps in preprocessing. NBIs were calculated for the satellite images. The pastures and the surrounding grassland area which were investigated *in situ*, were manually defined in the satellite images. Subsequently, biomass and vegetation cover have been estimated for all pixels belonging to the investigated areas. The performance of the regionalization was validated by comparing the measured values with the predicted values at the respective site where the samplings were performed.

6.3 Results

6.3.1 Feature selection and predictive importance of NBIs

Regarding the prediction of vegetation cover, the NBIs selected for final RF training in the feature selection step indicated that important predictor variables were those including at least one band in the red edge or those calculated from channels between the green and the red part of the electromagnetic radiation (Fig. 6.3). The NBIs selected for final training of AGB did not show such a clear pattern but a majority of the important NBIs were combinations of green bands and red edge bands (Fig. 6.4).

6.3.2 Accuracies of estimations based on hyper- and multispectral data

Using cross validation to estimate the accuracies of the models when faced with unknown data, the model using hyperspectral bands for predicting vegetation cover showed a high performance of $R^2 = 0.89$ and a RMSE of 9.26% (Tab. 6.5). The loss of explained variance was only 1% when multispectral NBIs were used instead of hyperspectral ones (Tab. 6.5). There were no relevant differences between the three sensors regarding the ability to predict vegetation cover.

The performance of the models to predict AGB was lower. Using the hyperspectral feature space, the cross validated R^2 was 0.31 and the RMSE was 20.71 g. As for vegetation cover, a resampling to multispectral bands of QB, RE and WV did not affect the performance of the resulting models. A resampling to WV even yielded a slightly higher R^2 of 0.35 compared to the hyperspectral model.

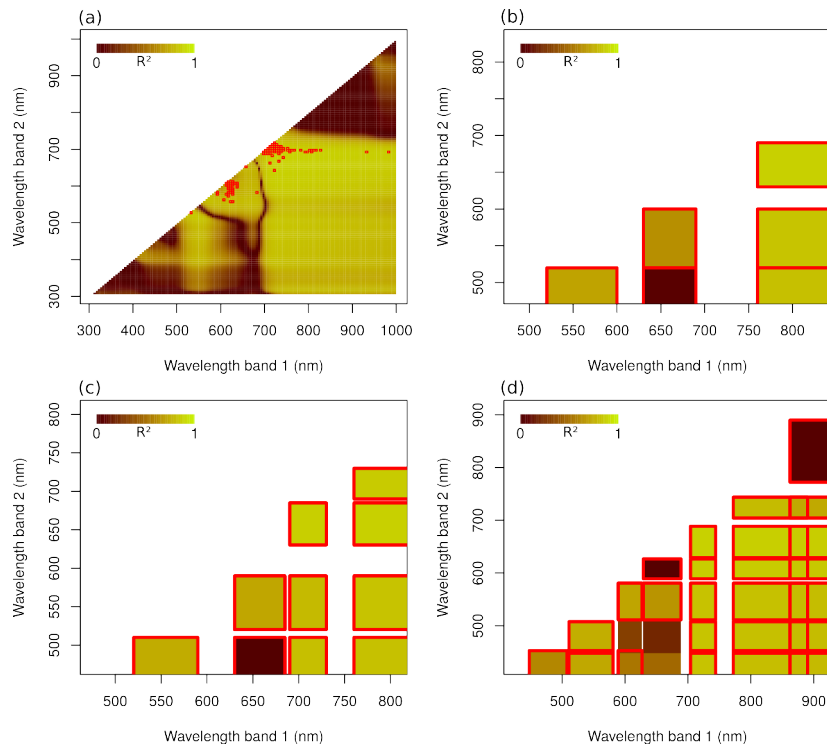


Figure 6.3: Correlograms depicting the R^2 values (color) of linear regressions between NBIs calculated from reflectance values at the x - and y -axes and vegetation cover. Optimal performing NBIs in the RF model revealed by recursive feature selection are marked by red squares. In the hyperspectral model only the 100 best performing NBIs are depicted. Models used all hyperspectral NBIs (a) as predictor variables or NBIs of bands resampled to QB (b), RE (c), WV (d).

Table 6.5: Cross validated results of the feature selection with RF training for the different models calculated in this study. Units of RMSE are % and g/m^2 respectively.

Model	No. of Variables	R^2	RMSE
Vegetation cover			
Hyperspectral	1350	0.89	9.26
Quickbird	7	0.88	9.68
Rapid Eye	11	0.88	9.65
World View-2	28	0.88	9.66
AGB			
Hyperspectral	3500	0.32	20.71
Quickbird	5	0.32	20.95
Rapid Eye	8	0.31	21.25
World View-2	30	0.35	21.28

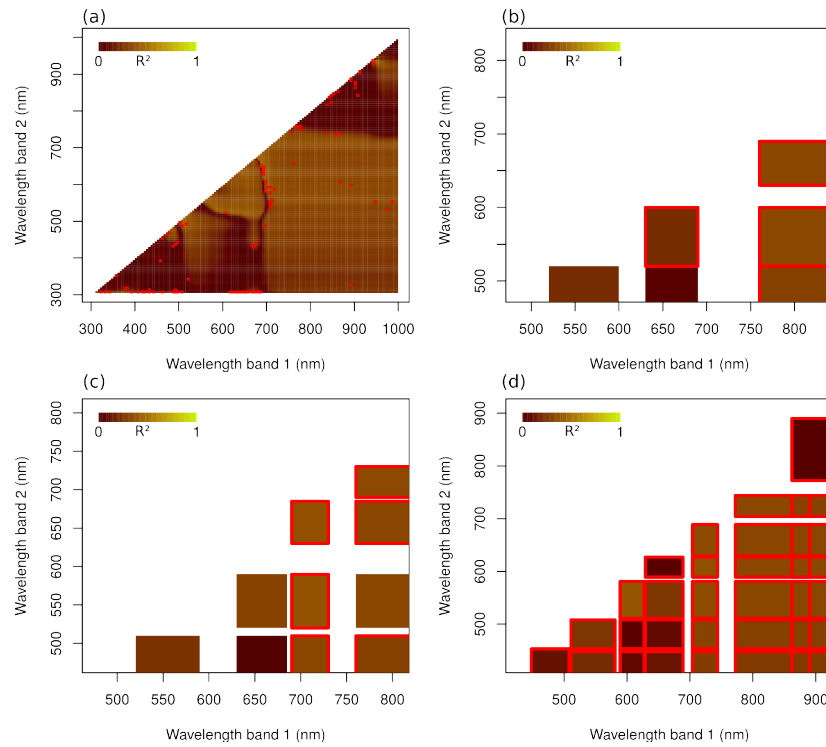


Figure 6.4: R^2 values of linear regressions between NBI and AGB represented by color. See Fig. 6.3 for further explanations.

6.3.3 Application to satellite data

Fig. 6.5 shows the results of the models applied to a sample satellite image. In this case the QB image from the site Qumahe was chosen since this location features a wide range of vegetation cover and AGB. The performance of the model applied to the satellite images decreased compared to the simulated images (Fig. 6.6, 6.7). On average, the performance of the vegetation cover models dropped from an overall $R^2 = 0.99$ using the simulated spectra to $R^2 = 0.74$.

While the performance using the simulated spectra did not change between the individual satellite sensors, the performance differed when applied to the satellite images. The vegetation cover samples which were located in QB images could be predicted with a R^2 of 0.48. The results showed a high performance for sites with low vegetation cover, whereas high vegetation cover sites (> 0.4) were underestimated by the model (Fig. 6.6 b). The samples located in RE scenes were predicted with a R^2 of 0.53 and the WV samples with $R^2 = 0.87$. The RE model results showed that sites with low vegetation cover could successfully be separated from sites with high vegetation cover (Fig. 6.6 d). However, it was lacking an accurate grading within the two clusters.

Regarding AGB, the overall R^2 decreased from 0.94 to 0.64 when satellite NBI values were used as predictor variables rather than the simulated satellite bands

from the *in situ* spectral measurements. The predicted AGB was not evaluated for the individual sensors since the sample size was too small to achieve meaningful results.

The distribution of the vegetation cover values at the different locations differed largely in that the highest vegetation cover values were observed in the eastern declivity while vegetation cover values in the arid western part were lowest (Fig. 6.8). The vegetation cover in the transition zone between Alpine steppes and Alpine meadows were highly variable. Here, values from below 10% up to 80% cover were observed. The distribution of biomass estimates generally followed that of the vegetation cover (Fig. 6.9). In general, the differences were less pronounced and higher ranges were observed. Especially in the transition zone, the distributions were complex with multiple maxima.

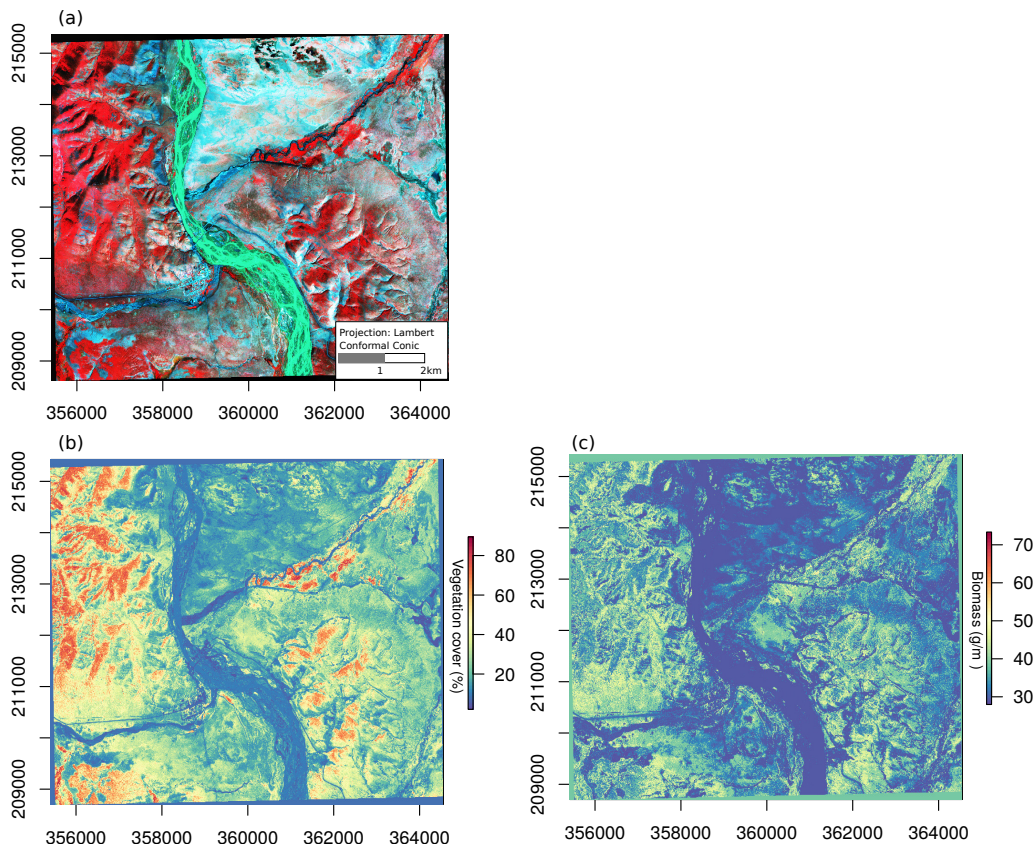


Figure 6.5: Example of the spatial predictions at the location "Qumahe". The QB scene is depicted as false color composite highlighting green vegetation (a). The albedo values in this image were used to predict vegetation cover (b) and AGB (c).

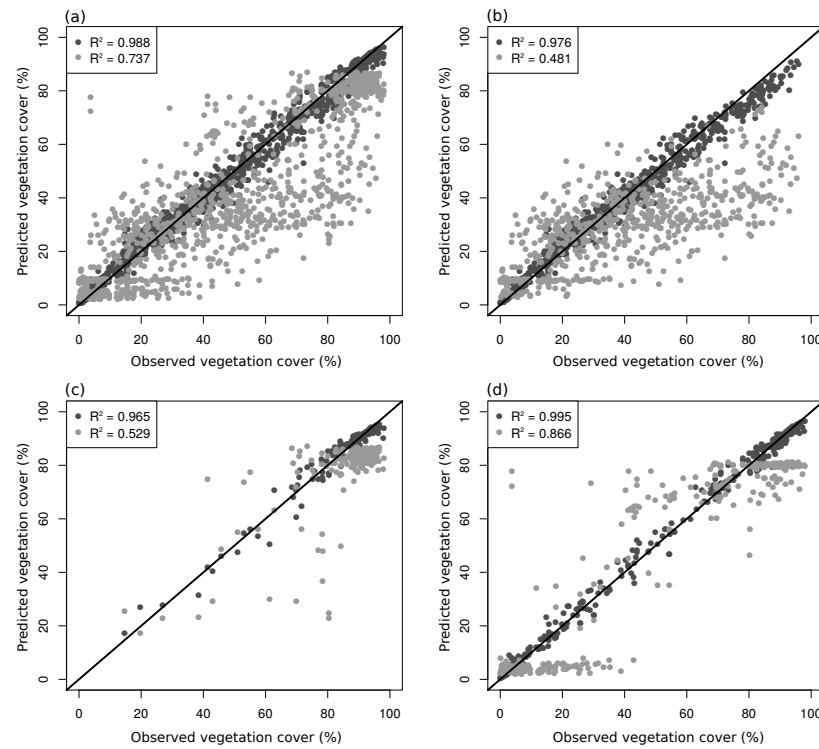


Figure 6.6: Validation of the vegetation cover models. (a) shows the predictions for all sensors, (b), (c), (d) focus on QB, RE and WV, respectively. Dark gray points represent the results from models using the simulated spectra and light gray points represent the results when the models are applied to the satellite images.

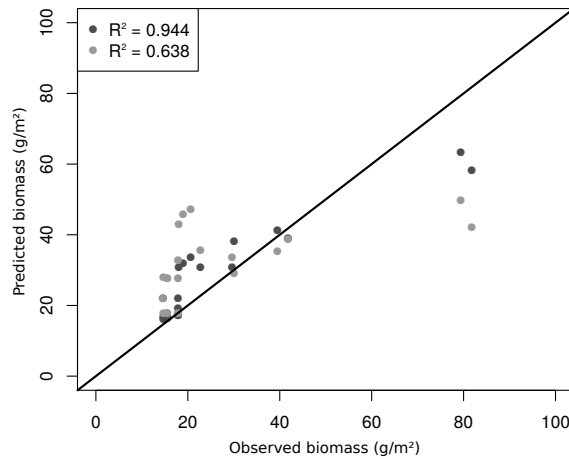


Figure 6.7: Validation of the AGB models. The predictions for all sensors are represented. See Fig. 6.6 for further explanation.

6.4 Discussion

Our approach based on NBIs and RF to predict vegetation cover and AGB as proxies for pasture degradation on the QTP yielded promising results. This highlights the great potential of machine learning algorithms to derive vegetation

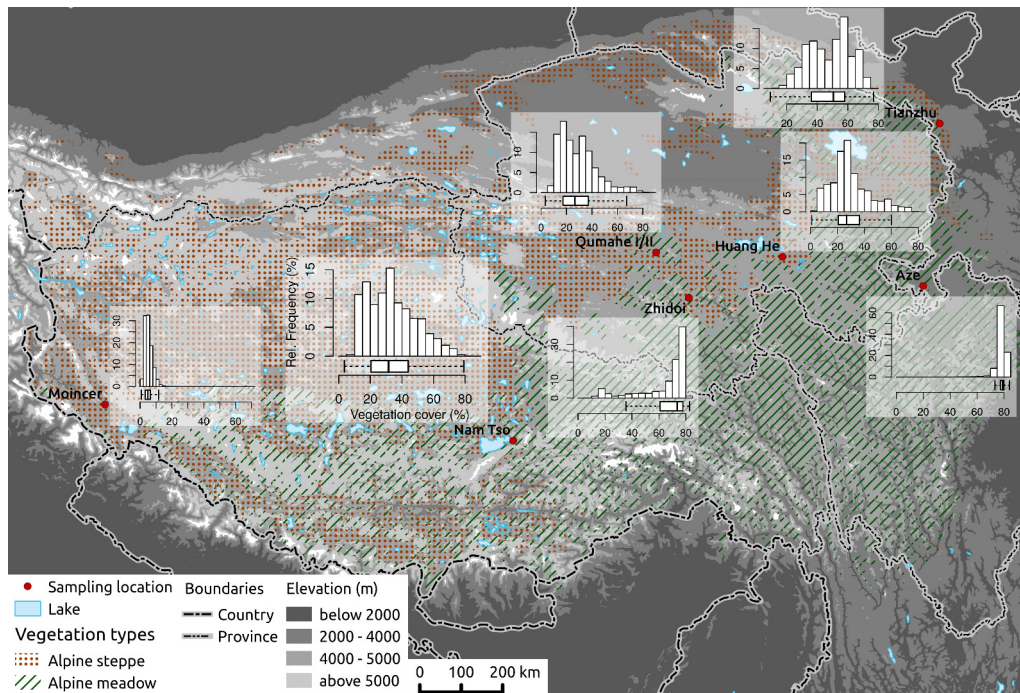


Figure 6.8: Distribution of vegetation cover values at the different locations where satellite data was available. For axis labels see the plot at Nam Tso location. Note that outliers are omitted in the boxplots; lines and boxes correspond to extreme values, the first quartile, third quartile, and the median.

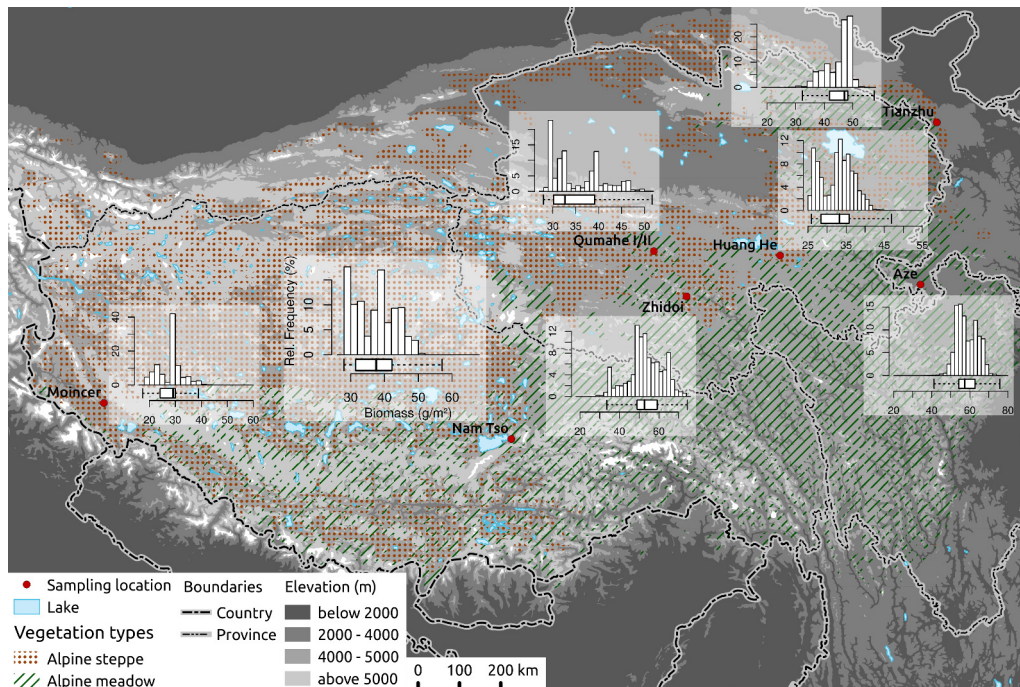


Figure 6.9: Distribution of biomass values at the different locations where satellite data was available. See Fig. 6.8 for further explanations.

parameters from remotely sensed datasets. This has already been shown regarding the estimation of e.g. the leave area index (Vuolo *et al.*, 2013) or chlorophyll

content (Verrelst *et al.*, 2012). Our results highlight the high ability of all tested satellite sensors to predict vegetation cover using RF as a modeling tool. Although, the large scale prediction of AGB was more afflicted with uncertainties than that of vegetation cover, the accuracies were higher than in previous studies only using vegetation indices as proxies for biomass (Yang *et al.*, 2009).

Resampling to multispectral bands only resulted in a slight decrease of the performance compared to the hyperspectral models. Therefore, the tested multispectral satellite sensors provide a sufficient spectral resolution to estimate vegetation cover and AGB. The usage of hyperspectral data as e.g. applied in Psomas *et al.* (2011) or Beerli *et al.* (2007) for AGB modeling is therefore not necessary. Consequently, multispectral datasets are preferred for the task of vegetation cover and AGB modeling on the QTP, because hyperspectral sensors have a lower data availability and higher computational effort.

Though the models showed a high performance using the simulated spectra of QB, RE, WV, the performance decreased when applied to the satellite images. The step to actually transfer models based on *in situ* measurements to satellite data was not performed in the majority of the studies aiming to reveal the potential of *in situ* spectral measurements to map pasture characteristics (e.g. Shen *et al.* (2008); Fava *et al.* (2010); Numata *et al.* (2008)). However, estimates about the loss of performance are important to assess the potential of *in situ* measurements in a regionalization of pasture degradation. The decrease in performance can be explained by a combination of several factors including the spatial scale of the satellite images, time gaps between measurement and image acquisition as well as small deviations between the measured site and the actual position in the image. Due to the larger scale of the satellite image, heterogeneity in vegetation cover and AGB or even patchiness within the extent of a pixel may affect the results. To overcome this problem, the measured area must be absolutely representative for the whole pixel in which they are located. However, this can not always be taken for granted, especially not in the alpine steppes where patchiness is partly a characteristic of the vegetation type. The issue of scale is clearly reflected by the data: WV with the best spatial resolution is closest to the performance of the simulated spectra and QB with a lower resolution showed the lowest performance. The issue of patchiness of the vegetation is reflected in the QB based model (Fig. 6.6 b): Since most of the sites in the QB model are from alpine steppe, sites with high vegetation cover are compensated when the size of a whole pixel is considered. Therefore, the model underestimated vegetation cover of sites which actually had a high cover. A truly direct comparison between the QB, RE and WV models is not possible since the images are from different locations, thus the sample size as well as the range of the observed vegetation cover and AGB is not the same for all images.

Nevertheless, one must consider that the issues of spatial scale, time gaps as well as small location errors affect the validation of the model. Since the model was calculated using spectra from the exact position where vegetation cover was determined and AGB was harvested, and also considering that there was no time gap between determination of the response variables and the measurement of the spectra, the regionalization might be even better than shown by the validation results.

Aside from the issue of scale, the data reflect a common drawback of the RF method. Since RF is an ensemble method that averages the predictions of several individual models, extreme values cannot be captured. This pattern becomes obvious in the validation of the WV based models (Fig. 6.6 d) where a saturation effect regarding vegetation cover $> 80\%$ is the consequence. However, since most of the area features a lower plant cover, this saturation will only have an effect in extremely moist areas at the eastern declivity (Lehnert *et al.*, 2015b).

Our estimations of vegetation cover revealed distinct patterns regarding the large scale differences of the plateau's vegetation and its small scale variability. Generally, estimates of vegetation cover are difficult to be compared to observations from literature since most studies report sums of species cover rather than overall cover values. If compared to other field studies in the western part, our model tended to underestimate the vegetation cover in the arid western part by approx. 5 - 10% (Miehe *et al.*, 2011a). In the transition zone, the estimations of our model were in good agreement to values published from field surveys (Babel *et al.*, 2014). The spatial pattern of biomass estimations were less pronounced than those for vegetation cover. This highlights that the biomass models were largely independent from the vegetation cover observed at the point of measurement. Thus, separate models have to be trained to estimate biomass on the QTP rather than to use simple transfer functions between vegetation cover - or even NDVI - and biomass (Shen *et al.*, 2008). Our biomass estimations were similar to other field observations (Yang *et al.*, 2009).

It could be shown that *in situ* measurements are well suited to establish reliable models that are able to predict vegetation cover and AGB based on NBIs. The advantage of the models which are calculated using measurements of different locations on the QTP is twofold. (1) They can be applied to a wide range of high resolution satellite sensors which is particularly important because of a lack of data availability on the QTP. (2) The approach allows estimating pasture degradation proxies at remote locations which may be partly outside of the investigated area. This can not be achieved if models would be calibrated by directly using the satellite data as predictors such as in e.g. Zha *et al.* (2003).

Field samplings were carried out at 18 locations on the QTP and regionalization bases on 7 satellite images. Though the locations were selected with regrad

to representativity for the respective area and a view to obtain a full gradient of vegetation coverage, it is of note that a direct transferability of the models to the entire plateau must be considered with care. It is of further note that due to the small sampling size of AGB, a leave-one-location-out cross-validation of the AGB models was not possible, therefore the transferrability of the model to "unknown" locations could not be assessed. In further studies, it would be desirable to extend the database of ground truth data in order to improve the significance of the AGB model.

The seasonality of the vegetation as well as long-term trends have been studied on a low spatial resolution (Jin *et al.*, 2013; Lehnert *et al.*, 2016; Sun *et al.*, 2013; Zhong *et al.*, 2010). Those studies show that there are pronounced seasonal changes in the phenology of the grasslands. However, the proxies for pasture degradation provided by our study give an instantaneous impression about the status of the pastures on a high spatial resolution. The assessment of trends and phenological changes on such a high spatial resolution is unfortunately not possible. Nevertheless, future satellite missions (e.g. sentinel) may provide more appropriate sources of data.

6.5 Conclusions

The aims of this study were (i) to test if hyperspectral data outperforms multispectral data when estimating vegetation cover and AGB on the QTP, and (ii) to assess the possibility to derive high spatial resolution maps of these pasture degradation proxies. It has been shown that *in situ* measurements from different sites are suitable to calculate models that predict pasture degradation proxies. In this context, RF has been shown to be a promising tool to deal with the high amount of predictor variables. Whilst vegetation cover could be estimated with high accuracy, the prediction of AGB was less accurate and is a challenging pasture degradation proxy for future studies.

Regrading our first aim, the spectral resolution of the three commonly used high resolution satellite sensors QuickBird, RapidEye and WorldView-2 provide sufficient information. Thus, hyperspectral data have no advantage compared to multispectral data in the scope of vegetation cover and AGB modeling.

Concerning the second aim it could be shown that though the simulated satellite bands using hyperspectral measurements as basis were equally well performing as the hyperspectral models, the performance decreases when actually applied to the satellite data. This emphasizes the issue of scale differences between the locally taken samples and the spatial resolution of the images. The mentioned issues in the regionalization process are not solely subject to this study but a

challenge that all studies aiming at transferring *in situ* measurements to a regional scale are faced with. Despite these uncertainties, the results highlight the potential of locally taken samples in combination with multispectral satellite data for the plateau wide regional mapping of pasture degradation proxies.

This study provides a first attempt to derive high resolution maps of pasture degradation proxies on the QTP which is urgently required to gather spatially explicit information on pasture degradation. The knowledge of pasture degradation is important for local people and politicians to maintain ecosystem services provided by the largest high mountain grassland ecosystem of the world.

Acknowledgments This study was conducted within the framework of the PaDeMoS- Project ("Pasture Degradation Monitoring System") and was funded by the German Federal Ministry of Education and Research (03G0808C). We thank Michael Göbel, Alexander Groos, Nele Meyer, Sebastian Semella, and Eva-Vanessa Wilzek for their help in the field and processing the large data sets.

Chapter 7

Automatic classification of Google Earth images for a larger scale monitoring of bush encroachment in South Africa

Annika Ludwig (1), Hanna Meyer (1), Thomas Nauß (1)

(1) Faculty of Geography, Philipps-University Marburg, Deutschhausstr. 10, 35037 Marburg, Germany

Published in
International Journal of Applied Earth Observation
and Geoinformation, 2016, 50, 89–94

Received 13 October 2015

Revised 12 January 2016

Accepted 8 March 2016

Available online 25 March 2016

Published August 2016

7 Automatic classification of Google Earth images for a larger scale monitoring of bush encroachment in South Africa

Annika Ludwig, Hanna Meyer, Thomas Nauß

Abstract

Bush encroachment of savannas and grasslands is a common form of land degradation in the rangelands of South Africa. To assess the carrying capacity of the land and to understand underlying processes of bush encroachment, continuous monitoring of this phenomenon is needed. The aim of this study is to provide training sites for satellite-based monitoring of bush encroachment in South Africa on a medium spatial resolution satellite sensor (e.g. MODIS or Landsat) scale. Since field surveys are time consuming and of limited spatial extent, the satellite based creation of training sites using Google Earth images is intended. Training pixels for woody vegetation and non woody land cover were manually digitized from 50 sample Google Earth images. A Random Forests model was trained to delineate woody from non woody pixels. The results indicate a high performance of the model (AUC=0.97). The model was applied to a further 500 Google Earth images with a spatial extent of 250 x 250m. The classified images form the database of training sites which can be used for larger scale monitoring of bush encroachment in South Africa.

Keywords Bush encroachment; Environmental Monitoring; Google Earth; Random Forests; Rangelands; South Africa

7.1 Introduction

Bush encroachment of arid and semi-arid savannas and grasslands is seen as a common form of land degradation in the rangelands of South Africa. Bush encroachment is defined as the suppression of palatable grasses and herbs by woody vegetation which are palatable to browsers but not eaten by the majority of domestic livestock (Ward, 2005). The negative economic consequences are enormous since grass dependent livestock represents the main income for many local farmers who are now faced with a reduced carrying capacity of their land (Ward, 2005). Therefore, a monitoring of bush encroachment is needed for several reasons: it allows farmers to identify locations with upcoming bushes giving them a tool for management and allows them to assess the current carrying capacity of their land. It also serves scientists as a baseline to reveal the yet unknown causes of bush encroachment.

Since spatially extensive field surveys are very cost extensive, remote sensing represents the only way to meet the demand of a high resolution, quasi-continuous and area wide monitoring. There are a number of case studies aiming at monitor bush encroachment in South Africa using remote sensing data (Hudak and Wessman, 1998; Hudak and Wessman, 2001; Munyati *et al.*, 2011; Symeonakis and Higginbottom, 2014). What these case studies have in common is that the spatial expansion of the product was limited by the availability of the ground truths. A variety of field surveys which provide ground truths were conducted in the South African rangelands on a local scale (Dreber *et al.*, 2014; Skarpe, 1991; Wiegand *et al.*, 2005; Britz and Ward, 2007; Roques *et al.*, 2001; Buitenwerf *et al.*, 2012). Though field surveys undoubtedly represent the most accurate way to identify training sites, they rarely match the spatial extent of medium resolution satellite systems (e.g. MODIS or Landsat) which are more suitable for operational monitorings. Therefore, satellite-based training sites with a spatial extent large enough to match at least one pixel of medium resolution satellite systems are needed to serve as ground truth in larger scale monitoring of bush encroachment.

There are a number of high resolution satellite products available which allow an accurate classification into woody vegetation and non woody land cover to provide ground truths for larger scale monitoring. Gessner *et al.* (2013) classified Quickbird images from Namibia into woody and non woody vegetation and used these classifications as ground truth for Landsat based estimations for the percentage of woody vegetation. WorldView images also provide a high resolution which enable delineation of woody from non woody land cover. Though these images have not yet been explicitly applied in upscaling chains aiming at the estimation of woody vegetation, they have been successfully used as training sites

for analysing grass cover based on WorldView-2 data on a Landsat and MODIS scale Lehnert *et al.* (2015b). The drawback of high resolution images like WorldView and Quickbird are their relatively high costs, often a limiting factor for the accessibility of images for research projects. In contrast, Google Earth images are free of charge and offer a high spatial resolution which makes them perfectly suited for generating training sites. Though Google Earth images are often used as ancillary data source to digitize training sites, they are rarely used as a direct data source for land cover classifications (Aher *et al.*, 2014; Almeer, 2012; Hu *et al.*, 2013). This might be due to two central challenges: Firstly, Google Earth images are only available in RGB bands and feature no infrared channel which is commonly used for classifying vegetation. Secondly, they are only available at fixed dates which differ between locations. While the first drawback might be overcome by use of a visible vegetation index (VVI), the second issue is more challenging. However, as machine learning algorithms (e.g. Random Forests) are more extensively used an increasing number of monitoring strategies can build more general models between reflectance and percentage of a land class (e.g. Gessner *et al.*, 2013) rather than estimating woody vegetation from single scenes only. Following such approaches where training sites are taken from multiple scenes, the acquisition date of the Google image is less important so long as a Landsat/MODIS image is available for the date of the Google image.

This study aims to provide training sites for an upcoming satellite-based monitoring of bush encroachment on a medium spatial resolution scale (e.g. MODIS or Landsat) in South African savannas and grasslands. To pursue this target we use Google Earth images and Random Forests to automatically delineate woody vegetation from non woody land cover. The classified images will form the database of training sites for the upcoming larger-scale monitoring.

7.2 Methods

The work flow of this study (Fig. 7.1) first requires example 50 Google Earth RGB images as baseline. From these images (i) training pixels for woody and non woody areas were manually digitized and (ii) derived predictor variables were calculated from the Google Earth RGB images. A Random Forests model was then trained to delineate woody from non woody pixels using a subset of the training pixels. The model was validated using the hold out samples. In a further step, the model was applied to 500 randomly chosen Google Earth images. The reliability of each classification was assessed using the predicted probabilities for woody vegetation. Using the reliable classifications only, the predicted percentage of woody vegetation will serve as input for the upcoming larger scale monitoring of bush encroachment. The following sections describe these steps in detail.

All steps of modeling and analysis were performed using the R environment for statistical computing (R Core Team, 2013). The caret package (Kuhn, 2014a) as a wrapper package for machine learning algorithms implemented in R was applied for model tuning, training and prediction.

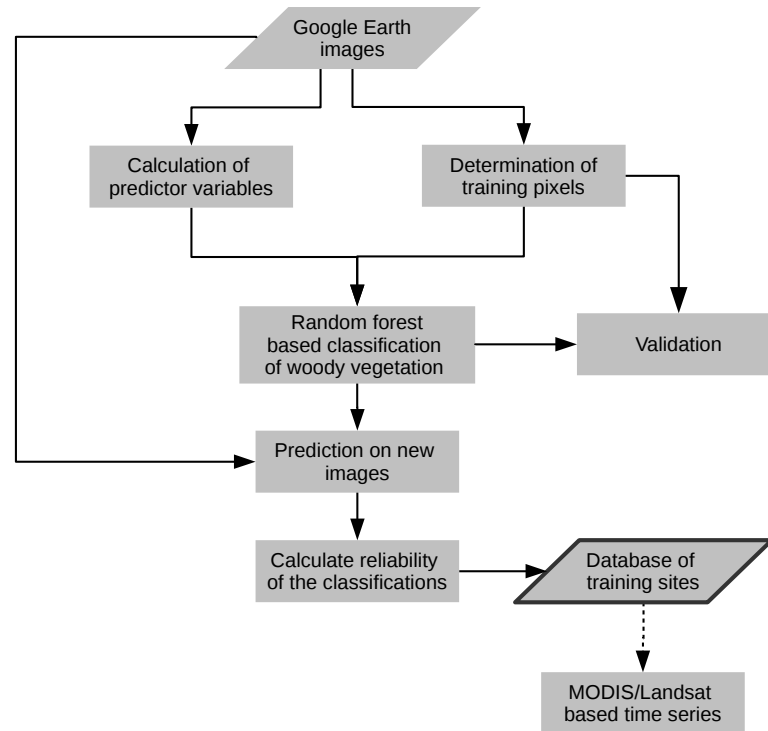


Figure 7.1: *Overview of the processing flow.*

7.2.1 Study area

The areas of interest in this study are the savanna and grassland biomes of South Africa, Lesotho and Swaziland (Fig. 7.2). Both biomes, savannas and grasslands, are characterized by a mixture of grasses and sparse trees or bushes and affected by the problem of bush encroachment. See Mucina and Rutherford (2006) for further description on the vegetation of South Africa.

The determination of the study area was done on the basis of the biome classification of Mucina and Rutherford (2006). Only areas classified as savanna or grassland were taken into account. In a second step, all anthropogenic areas as defined by the MODIS land cover product were masked so that only savanna and grassland areas were considered for further analysis.

7.2.2 Data and variables

50 Google Earth images were downloaded at randomly chosen locations within the study area. Each image had a spatial extent of 250 x 250m which corresponds,

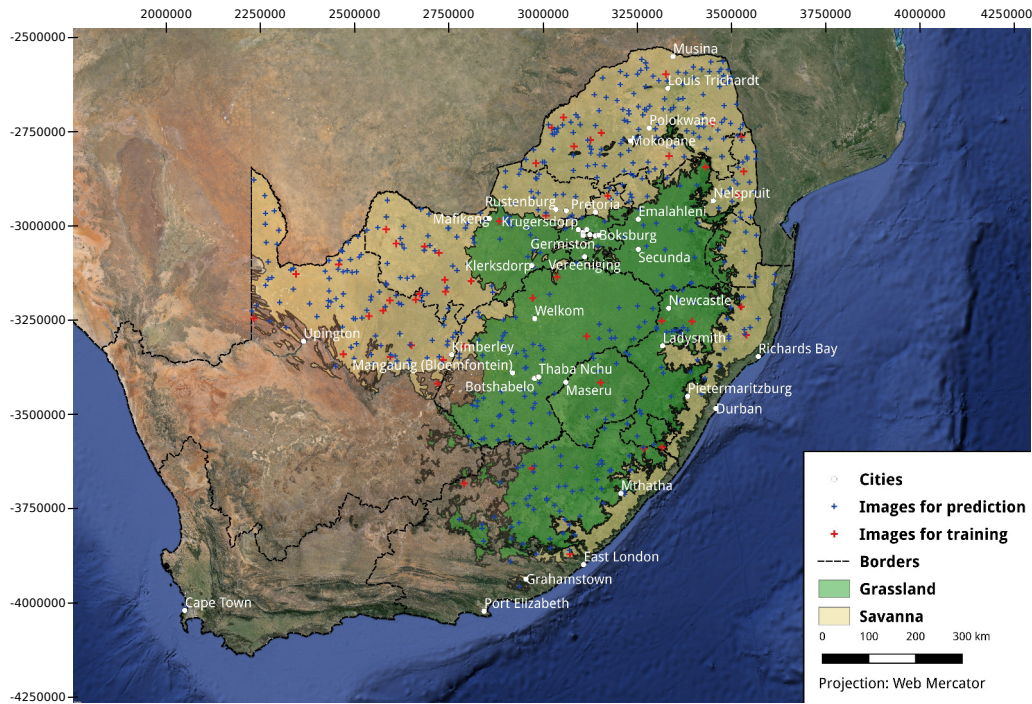


Figure 7.2: Map of the study area including the location of training images as well as images used for prediction. Only savanna and grassland biomes were included in the analysis.

or exceeds the size of a pixel from medium spatial resolution satellite sensors (e.g. MODIS or Landsat). The images were downloaded as georeferenced RGB images using the `gmap` function from the `dismo` package in R (Hijmans *et al.*, 2015). The highest available spatial resolution for each respective image was used, corresponding to a pixel size of approximately 30 by 30cm. For all images, the RGB values as well as the HSV values and a visible vegetation index (VVI, described in e.g. Joseph and Devadas, 2015) were used as predictor variables. The vegetation index takes advantage of the spectral properties of vegetation in the visible spectrum of light to distinguish between vegetated and non-vegetated surfaces. Additionally, texture measures were included by calculating the mean and standard deviation (sd) values in a 3x3 environment of all 7 spectral variables. The two-level variable "Biome" was further included to account for differing land cover characteristics between grasslands and savannas. In total, 22 predictor variables were used.

Training sites for woody vegetation and non woody land cover were manually digitized from the 50 Google Earth images (in total 220507 pixels for woody vegetation and 283289 pixels for non woody land cover). The values of the predictor variables at each of the training pixels were extracted from the raster data.

7.2.3 Random Forests classification

Random Forests was used to create a model which delineates woody vegetation from non woody land cover. The Random Forests (RF) algorithm of Breiman (2001a) is based on the concept of regression and classification trees. Random Forests repeatedly builds trees from random samples of the training data. In classification models, the class which is most often predicted from the individual trees is taken as the final estimate. To overcome the correlation between trees, a subset of predictors (mtry) is randomly selected at each split. The best predictor from the random subset is used at the respective split to partition the data. Random Forests was chosen for several reasons: (i) because it showed good performance in land cover classification applications (Gislason *et al.*, 2006), (ii) it performs as an intuitive classifier which can handle different types of variables and associated ranges of their values, thus no standardization is required. Further, Random Forests allow to estimate variable importances which allow interpretation despite being a black box.

For the modeling process, the set of training pixels was randomly split into a training (25%) and a test (75%) data set using stratified sampling which accounts for the distribution of each class. Using only the training data set, mtry was tuned between 2 and 22 using a stratified 10 fold cross validation. The number of trees (ntree) was set to 500 since no increase of performance could be observed from 500 trees onwards. Further, since Random Forests is robust to non informative variables (Kuhn and Johnson, 2013a), there was no need for further feature selection. Models were fitted by repeatedly leaving one of the folds out. The performance of each model was determined by predicting the respective held-out fold. The performance metrics from the hold-out iterations were averaged to the overall model performance for the respective mtry value. Receiver operating characteristics (ROC) was used as performance metric to find the optimal model.

7.2.4 Validation

The hold out testing samples of the training pixels were used for a completely independent model validation. The metrics used for validation were receiver operating characteristics (ROC), probability of detection (POD), probability of false detection (POFD) and false alarm ratio (FAR). ROC curves as metric for model selection describe a model's performance independently of the probability threshold which separates woody vegetation pixels from non-woody pixels (Fawcett, 2006; Hamel, 2009). A ROC curve is constructed by iterating through multiple of probability thresholds to classify the probabilities for woody vegetation as output of the Random Forests model into woody vegetation and non woody land cover. At each threshold the data set is classified into woody vegeta-

tion and non woody land cover and the confusion matrix is calculated using the hold out samples for model validation. For each confusion matrix, the true positive rate and the false positive rate is calculated (see e.g. Hamel, 2009). The ROC curve is then drawn by plotting the true positive rate of each threshold against the corresponding false positive rate. In contrast to the confusion metrics based on the classified data, there is consequently not only one scalar value rating the model but several scalar values, each for one threshold. Therefore, ROC ensures that a model which assigns e.g. a probability value of 0.49 for a woody vegetation pixel can be interpreted as a better model than a model with a probability value of 0.1. If only confusion metrics were considered, then both models would be rated as being equal (i.e. non-woody) because both probabilities are under the default threshold of 0.5. This characteristic makes the ROC curve suitable and intuitive for classification model validation and comparison. The area under the ROC curve (AUC) gives the overall model performance and can be interpreted as the probability that woody vegetation and non woody land cover are correctly separated by the model.

In addition to the ROC values categorical verification scores were calculated from confusion matrices (Tab. 7.1) showing agreement and disagreement between predicted and observed woody vegetation based on the default 0.5 probability threshold (Tab. 7.2). The probability of detection (POD) gives the percentage of woody vegetation pixels that were correctly identified as woody vegetation by the model. The probability of false detection (POFD) gives the proportion of non woody pixels that were falsely classified as woody vegetation by the model. FAR gives the proportion of incorrectly predicted woody vegetation pixels from all pixels that were predicted as woody vegetation.

Table 7.1: *Confusion matrix as baseline for the calculation of verification scores used for the validation of the prediction model.*

	Observation		
		Woody	Non Woody
	Prediction	Woody	Non Woody
	Woody	True positives (TP)	False positives (FP)
	Non Woody	False negatives (FN)	True negatives (TN)

7.2.5 Prediction on new Google Earth images

To build the database of training sites, 500 randomly chosen sites within the study area were selected and the corresponding Google Earth images were downloaded. The predictor variables were calculated for the 500 images as described above and the fitted models were used to classify the images. Random selection of images leads to the complication that some images did not appear to be representative for the study area due to unusual patterns or the strongly deviating

Table 7.2: *Cross tabulation based validation metrics*

Metric	Formula	Range	Optimal value
Probability of detection	$POD = \frac{TP}{TP+FN}$	0-1	1
Probability of false detection	$POFD = \frac{FP}{FP+TN}$	0-1	0
False alarm ratio	$FAR = \frac{FP}{TP+FP}$	0-1	0

brightness of the images. To overcome this issue, the reliability of the image classifications was calculated on the basis of the probabilities for both classes which were predicted by the Random Forests model. For this purpose, the percentage of pixels which had a probability larger than 75% for either woody or non woody was calculated. From the 500 images, only those which had a higher reliability than the first quartile of all 500 images were finally taken for the database. The image acquisition date was obtained by loading the randomly chosen locations into Google Earth and manually reading the date of the images at the highest zoom level.

7.3 Results and Discussion

7.3.1 Model performance and variable importance

The revealed optimal value for mtry during model training was 10 which corresponds to an cross validated AUC of 0.996 indicating that there is an excellent fit of the training data. Applied on the independent test data, the model still showed a very high performance (Tab. 7.3, 7.4) with an AUC value of 0.974. 97% of the woody vegetation pixels in the test data set were correctly identified as woody vegetation by the model (Tab. 7.4). The probability that a non woody pixel was falsely classified as woody was very low (2%). These results show that using representative pixels for model training, woody vegetation and non woody land cover can be successfully delineated on the basis of Google Earth RGB images and Random Forests.

The variable importance of the Random Forests model (Fig. 7.3) calculated according to Liaw and Wiener (2002) revealed a significance of the VVI as predictor variable in the model. Its 3 x 3 standard deviation and mean values were the most important variables in the model. In addition, the pixel based VVI was among the 10 most important variables (out of 22). RGB channels as well as hue

and saturation were also identified as important variables, just like the biome variable.

Table 7.3: *Contingency table of the test data set.*

	Woody	Non Woody
Woody	159763	4002
Non Woody	5617	208464

Table 7.4: *Performance of the Random Forests model calculated using the test data set.*

AUC	POD	POFD	FAR
0.974	0.966	0.019	0.024

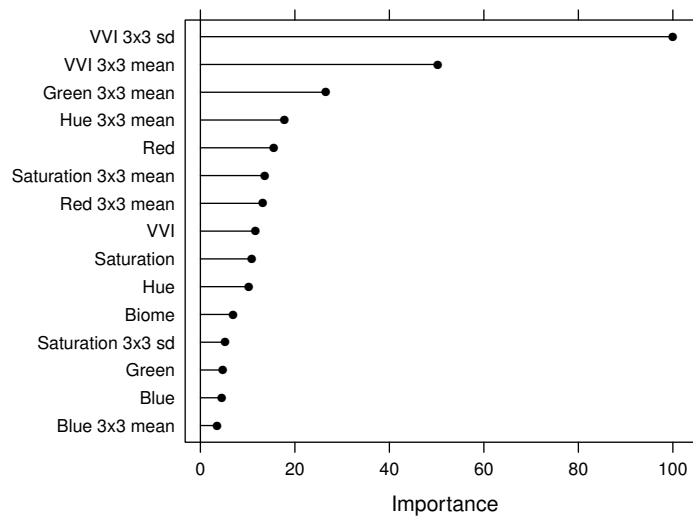


Figure 7.3: *Variable importance of 15 highest ranked predictor variables in the Random Forests model. Importance values were scaled to have a maximum value of 100.*

7.3.2 Database of training sites

The model was based on a selection of representative pixels for woody vegetation and non woody land cover. For the application on the new images, the model was faced with an increased variability in the spectral characteristics of both, woody and non woody land cover. Therefore, the need for a reliability check was important to account for e.g. strong deviations which were not trained by the model. Fig. 7.4 shows that the majority of the images were classified with high reliability. The first quartile of reliability was 70%. Images with a reliability lower than this threshold were rejected from the database. Using this threshold, 367 images from the initial 500 passed the reliability check. 133 images could not be classified with sufficient reliability. This is most likely due to structures

of land cover which were not captured in the manually digitized training pixels. An extension of the training pixels might therefore be meaningful but can only be achieved by field surveys since some patterns are visually hard to identify in the Google Earth images.

Fig. 7.5 shows three examples of different Google Earth images. The first one shows a clear savanna pattern: a slightly vegetated ground with patches of bushes and trees. This pattern was easily predicted by the model since the sandy ground clearly contrasts with the dark color of the woody vegetation. Accordingly, the model could classify this image with a high reliability resulting from very high probabilities for the corresponding woody or non woody class (Fig. 7.5, Tab. 7.5). The second example, from a grassland site, featured a high uncertainty in the upper part of the image. This uncertainty is reasonable, since it is even hard to rate by visual analysis of the satellite images whether these are small woody shrubs or dense grasses. Though the classification came up with a acceptable result, this image fails the reliability check, since only 64% of the pixels had probabilities larger than 75% for the corresponding land cover class. The third example, which is also from a grassland site, features a pattern of shrubs with different spatial extents in an area with medium ground cover by grasses and non woody plants. Compared to the first example, there is less contrast in the RGB between woody and non woody vegetation. The model reflects this pattern giving a high percentage of pixels with probabilities for woody vegetation around 50%, close to the threshold for classification as either woody or non woody. However, with a total reliability of 74%, this image was rated as accurate enough for the database. Overall, the images from sandy savannas could be classified with highest reliability due to high contrasts between woody and non woody land cover in the visible channels.

The final database (see excerpt in Tab. 7.5) includes further information with respect to a larger scale monitoring. The date of the image acquisition, the exact coordinates of the image, the resulting percentage of woody vegetation and the reliability of the classification. The number of images in the database was initially limited to 500 images due to the effort of manually reading the date of the images. However, the database can be extended at any time.

Table 7.5: *Excerpt of the training site database for the three example images shown in Fig. 7.5.*

Image	Date	Location	Woody (%)	Reliability (%)
Example 1	2014/01/20	23.6215, -26.0288	44.7	86.5
Example 2	2014/10/26	24.7972, -28.3681	44.2	63.9
Example 3	2015/06/07	26.1853, -29.8878	39.5	74.2

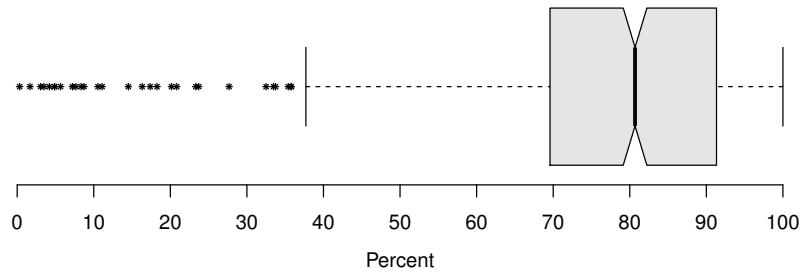


Figure 7.4: *Reliability of 500 Google Earth images. The first quartile (70%) was used as the threshold to select images for the database.*

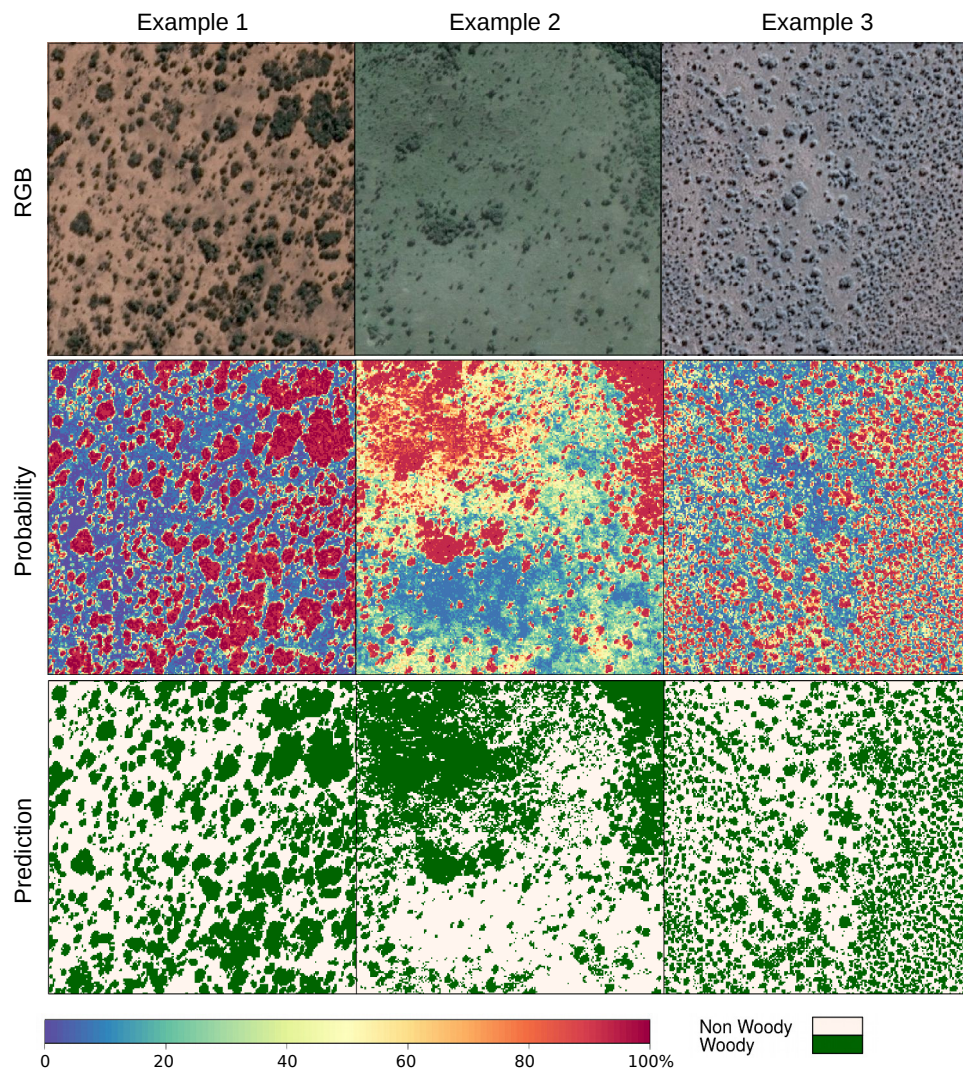


Figure 7.5: *Three example RGB images, the predicted probabilities for woody vegetation and the corresponding classification results.*

7.4 Conclusions

A method to develop low cost training sites for bush encroachment monitoring was presented. The study showed the capability of Google Earth images and Random Forests for automatic delineation of woody vegetation from non woody land cover. In total, 367 images were reliable enough for the database, covering a wide range of savanna and grassland in South Africa. The results are of significance for upcoming studies and area wide monitoring of bush encroachment. Since the extent of one image matches or exceeds the resolution of medium spatial resolution satellite sensors (e.g. MODIS or Landsat), the results for woody vegetation in the images can be used as training sites for a monitoring. Such monitoring could be based on a similar method as applied in this study: satellite data taken from the same date as the respective Google Earth images can be used together with the database from this study to train a model (using e.g. Random Forests) with the aim to predict bush percentage on the overall satellite image. In this way, a time series of bush encroachment for the savannas and grasslands of South Africa can be calculated. The database is available from the authors on request.

Acknowledgments This work was financially supported by the Federal Ministry of Education and Research (BMBF) within the IDESSA project which is part of the SPACES-program (Science Partnership for the Assessment of Complex Earth System processes).

Chapter 8

Spatio-temporal interpolation of soil water, temperature, and electrical conductivity in 3D + T: The Cook Agronomy Farm data set

Caley K. Gasch (1), Tomislav Hengl (2), Benedikt Gräler (3), Hanna
Meyer (4), Troy S. Magney (5), David Joseph Brown (1)

- (1) Department of Crop and Soil Sciences, Washington State University, USA
- (2) ISRIC — World Soil Information/Wageningen University and Research, The Netherlands
- (3) Institute of Geoinformatics, University of Münster, Germany
- (4) Department of Geography/Environmental Informatics, Philipps-Universität Marburg,
Germany
- (5) College of Natural Resources, University of Idaho, USA

Published in
Spatial Statistics, 2015, 14 (A), 70–90

Received 1 November 2014

Accepted 1 April 2015

Available online 29 April 2015

Published November 2015

8 Spatio-temporal interpolation of soil water, temperature, and electrical conductivity in 3D+T: the Cook Agronomy Farm data set

Caley Gasch, Tomislav Hengl, Benedikt Gräler, Hanna Meyer, Troy Magney, JDavid Brown

Abstract

The paper describes a framework for modeling dynamic soil properties in 3-dimensions and time (3D+T) using soil data collected with automated sensor networks as a case study. Two approaches to geostatistical modeling and spatio-temporal predictions are described: (1) 3D+T predictive modeling using random forests algorithms, and (2) 3D+T kriging model after detrending the observations for depth-dependent seasonal effects. All the analyses used data from the Cook Agronomy Farm (37 ha), which includes hourly measurements of soil volumetric water content, temperature, and bulk electrical conductivity at 42 stations and five depths (0.3, 0.6, 0.9, 1.2, and 1.5 m), collected over five years. This data set also includes 2- and 3-dimensional, temporal, and spatio-temporal covariates covering the same area. The results of (strict) leave-one-station-out cross-validation indicate that both models accurately predicted soil temperature, while predictive power was lower for water content, and lowest for electrical conductivity. The kriging model explained 37 %, 96 %, and 18 % of the variability in water content, temperature, and electrical conductivity respectively versus 34 %, 93 %, and 5 % explained by the random forests model. A less rigorous simple cross-validation of the random forests model indicated improved predictive power when at least some data were available for each station, explaining 86 %, 97 %, and 88 % of the variability in water content, temperature, and electrical conductivity respectively. The high difference between the strict and simple cross-validation indicates high temporal auto-correlation of values at measurement stations. Temporal model components (i.e. day of the year and seasonal trends) explained most of the variability in observations in both models for all three variables. The seamless predictions of 3D+T data produced from this analysis can assist in understanding soil processes and how they change through a season, under different land management scenarios, and how they relate to other environmental processes.

Keywords Digital soil mapping; Random forests algorithm; Regression-kriging; Regionalization; Soil sensor network

8.1 Introduction

Comprehension of dynamic soil properties at the field scale requires measurements with high spatial and temporal resolution. Distributed sensor networks provide frequent *in situ* measurements of environmental properties at fixed locations, providing data in 2- or 3-dimensions and through time (Porter *et al.*, 2005; Pierce and Elliott, 2008). While sensor networks produce ample data for observing dynamic soil properties, data processing for inference and visualization become increasingly difficult as data dimensionality increases. Ideally, the end product should consist of seamless interpolations that accurately represent the spatial and temporal variability in the property of interest. These products can then be used for predictions at unobserved locations, they can be integrated into process models, and they can simply aid in visualization of soil properties through space and time.

Multiple approaches have been developed for spatial interpolation of soil properties and digital soil mapping, including:

1. multiple regression models based on the soil forming factors, terrain attributes, spatial coordinates, or derived principal components (McKenzie and Ryan, 1999);
2. smoothing (splines) and neighborhood-based functions (Mitas and Mitasova, 1999);
3. geostatistics, or kriging, and variations thereof (see overviews by McBratney *et al.* (2003) and Hengl (2009)).

Of these, regression-kriging (Odeh *et al.*, 1995; Hengl *et al.*, 2007), which combines a multiple regression model (a trend) with a spatial correlation model (a variogram) for the residuals, produces unbiased, continuous prediction surfaces. Regression-kriging has been adapted for soil mapping with great success, in part because of the flexibility in defining the trend model as a linear, non-linear, or tree-based relationship between the response and predictors. Furthermore, regression-kriging relies on the incorporation of auxiliary data, providing mechanistic support for the soil property predictions.

The widest application of regression-kriging in soil science has likely been for producing 2-dimensional (2D) maps (Hengl, 2009). However, soil data is often also collected at multiple depths, and geostatistical interpolation techniques can be expanded to represent soil predictions across both vertical and horizontal space (Malone *et al.*, 2009; Veronesi *et al.*, 2012). Global predictions of multiple soil properties obtained from 3-dimensional (3D) regression models were recently

showcased by Hengl *et al.* (2014a). Here, spline functions define the vertical trend (depth) within the regression model, while horizontal trends are defined by covariate grids. These approaches are sufficient for understanding static soil properties across 2- and 3D space; however, modeling dynamic soil properties requires expansion of the geostatistical model to incorporate correlation in data through time (Heuvelink and Webster, 2001; Kyriakidis and Journel, 1999). Addition of temporal and/or spatio-temporal predictors can assist in explaining temporal variation in a response variable, but fitting a variogram model in 2D and time (2D+T) poses additional challenges (summarized by Heuvelink and Webster, 2001). Specifically, time exists in only one dimension and has a directional component, while spatial properties might be correlated in vertical, horizontal, or 3D directions. The easiest solution for approximating dependence across both space and time is based on anisotropy scalings, which relate horizontal distances to distance in depth and temporal separation.

Modeling 2D+T data has successfully been implemented for predicting soil water from repeated field-wide measurements obtained with time-domain reflectometry, directly with ordinary kriging (Huisman *et al.*, 2003), and with the incorporation of estimated daily evapotranspiration (Jost *et al.*, 2005) and net precipitation (Snepevangers *et al.*, 2003) as covariates. More recently, daily air temperature predictions have been produced from spatio-temporal interpolation models of weather station data at the regional (Hengl *et al.*, 2012) and global (Kilibarda *et al.*, 2014) scales. These models combine spatial covariates (terrain attributes) and spatio-temporal covariates (remotely sensed daily land surface temperature) in the trend model to explain greater than 70% of variation in weather station observations.

Previously, 3D+T data has been analyzed in a spatio-temporal context, wherein interpolations produce predictions in a slice-wise manner (i.e. by depth or by time point) (Bárdossy and Lehmann, 1998; Wang *et al.*, 2001; Wilson *et al.*, 2003). To our knowledge geostatistical methods have not yet been expanded to produce predictions from data collected in 3D and time (3D+T). This may be, in part, due to the rarity of quality 3D+T data. With each added dimension, the number of observations required for accurate interpolation increases, as does the need for ancillary (covariate) data if a regression-kriging model is applied. Theoretically, adapting the existing regression-kriging framework for predicting in 3D+T can follow the same mathematical logic as the models that scale up from 2D to 3D or 2D to 2D+T (Heuvelink and Webster, 2001). In that context, existing geostatistical tools for interpolating spatio-temporal data can also assist in modeling 3D+T data (Pebesma, 2012; Pebesma and Gräler, 2013).

In this paper, we demonstrate two approaches for interpolating 3D+T soil water, temperature, and electrical conductivity data (collected from a distributed

soil sensor network) at the field-scale: one that is based on using random forests algorithms, and one that is based on spatio-temporal kriging. The kriging model uses different dependence structures (i.e. variogram models) for horizontal, vertical, and temporal components, which are then combined using concepts from 2D+T geostatistics. The models were motivated by the existing geostatistical frameworks and incorporate spatial, temporal, and spatio-temporal covariates. We present the implementation of the models, accuracy assessments, visualization and applications of model output, and future directions for improvement with a long term objective to develop robust 3D+T models for mapping soil data that has been collected with high spatial and temporal resolution.

8.2 Materials and methods

8.2.1 The Cook Agronomy Farm data set

The R.J. Cook Agronomy Farm is a Long-Term Agroecosystem Research Site operated by Washington State University, located near Pullman, Washington, USA (46°47'N, 117°5'W; Figure 8.1). The farm is 37 ha, stationed in the hilly Palouse region, which receives an annual average of 550 mm of precipitation (Western Regional Climate Center, 2013), primarily as rain and snow in November through May. Soils are deep silt loams formed on loess hills; clay silt loam horizons commonly occur at variable depths (Natural Resource Conservation Service (NRCS), 2013). Farming practices at Cook Agronomy Farm are representative of regional dryland annual cropping systems (direct-seeded cereal grains and legume crops).

At 42 locations (stations), five 5TE sensors (Decagon Devices, Inc., Pullman, Washington) were installed at 0.3, 0.6, 0.9, 1.2, and 1.5 m depths. Locations were chosen from an existing non-aligned systematic grid and stratified across landscape units to represent the variability in terrain of Cook Agronomy Farm (Figure 8.1). Every hour, the 5TE sensors measure:

1. volumetric water content, (m^3/m^3),
2. temperature, ($^{\circ}\text{C}$),
3. and bulk electrical conductivity, (dS/m).

Data are stored on Em50R data loggers (Decagon Devices, Inc., Pullman, Washington), which are buried to allow data collection regardless of farm operations (seeding, spraying, and harvest). The sensor network has been in operation since 2009. For the purpose of this article, hourly sensor data was aggregated to

daily averages and all plots and statistical modeling refers to daily values. Sensor data collected for three years at one station and all five depths is illustrated in Figure 8.2, and hexbin plots (Sarkar, 2008; Carr, 2014) illustrate the distribution of all observations of all three variables across depth in Figure 8.3. Please note that absolute values of sensor readings require further correction for accurate interpretation. Thus, interpretation of the presented readings should focus on the observed relative changes.

In addition to the sensor readings, this data set contains spatial and temporal regression covariates either at 10 m resolution, or as a temporal measurement that is assigned to all possible locations in the area of interest at a given time step (hereafter, spatially constant). Dimensionality of the covariates differs: some covariates are available only in horizontal space (elevation, wetness index, vegetation images), some covariates are available as 3D images (soil properties) and some are available either in time (daily temperatures and rainfall from the nearest meteorological station) or spacetime (cropping identity). The covariates used for modeling water content, temperature, and electrical conductivity are described in Table 8.1. Note that only the response variables (sensor readings) exist in 3D+T, while the predictor variables are a combination of 2D, 3D, 2D+T, and temporal covariates.

The SAGA wetness index, a modification of the topographic wetness index (Beven and Kirkby, 1979), was derived from the digital elevation model (DEM) using the RSAGA package (Brenning, 2013) for R (R Core Team, 2014). A total of 11 Level 3A RapidEye images satellite images acquired between 2011 and 2013 were used to incorporate vegetation patterns on Cook Agronomy Farm. Images were pre-processed exactly as in Eitel *et al.* (2011). Following image pre-processing, spectral bands (near infrared—NIR and red-edge—RE) were mathematically converted into the Normalized Difference Red-Edge Index (Barnes *et al.*, 2000):

$$NDRE = \frac{NIR - RE}{NIR + RE} \quad (8.1)$$

The RE region of the electromagnetic spectrum has been shown to be superior to red (as used in the Normalized Difference Vegetation Index, or NDVI, Tucker, 1979) for mapping variations in plant chlorophyll and nitrogen content (Carter and Knapp, 2001; Lichtenthaler and Wellburn, 1983; Eitel *et al.*, 2007, 2008, 2009). The images were aggregated to produce one NDRE mean grid and one NDRE standard deviation grid, which were resampled from a 5 m to a 10 m grid to align with other covariate grids.

The 3D maps for the occurrence of the Bt horizon, bulk density (g/cm³), and

soil pH were generated using 184 soil profiles distributed across Cook Agronomy Farm (Figure 8.1) using the **GSIF** package for automated soil mapping (Hengl *et al.*, 2014b). Soil profiles were described using the National Soil Survey Center NRCS USDA guidelines for soil profile description (National Soil Survey Center NRCS USDA, 2011). To make the maps, the presence or absence of a Bt horizon was interpolated using a logistic regression-kriging model and the DEM, apparent electrical conductivity grids, soil unit description map (Natural Resource Conservation Service (NRCS), 2013) and depth as covariates. Bulk density and soil pH were predicted with regression-kriging models with the DEM, wetness index, soil mapping units, apparent electrical conductivity grids, and depth as covariates.

The daily meteorological data (precipitation, minimum and maximum temperature) were obtained from a weather station located 8 km from the farm in Pullman, WA (Western Regional Climate Center, 2013). Daily precipitation was transformed to cumulative precipitation, which reverts to zero after a period of precipitation. Meteorological covariates are only available in the time domain (i.e. they are assumed to be spatially constant).

As the only 2D+T covariate we used the cropping system classification maps, which are available each year from 2006 through 2013. The crop identities include: barley, canola, garbanzo, lentil, pea, or wheat, each with either a spring or winter rotation.

All sensor observations and covariates were assembled into a spatio-temporal regression matrix, using the overlay functionality of the **spacetime** package in R (Pebesma, 2012). The resulting spatio-temporal regression matrix was very large — even though we only included measurements from 42 stations, the matrix contained close to a quarter million records (about four years of daily measurements at 42 locations and five depths i.e. $4 \times 365 \times 42 \times 5 = 306,600$ - missing data = 219,240 water content observations, 222,614 temperature observations, and 222,065 conductivity observations).

8.2.2 Conceptual foundation for 3D+T modeling

We model water content, soil temperature, and electrical conductivity as a spatio-temporal process of a continuous variable Z , where Z varies over space and time. The statistical model of such a process is typically composed of the sum of a trend and a stochastic residual (Burrough, 1998; Heuvelink *et al.*, 2012; Kilibarda *et al.*, 2014). In this case we begin with a 3D+T model of the form:

$$Z(x, y, d, t) = m(x, y, d, t) + \varepsilon'(x, y, d, t) + \varepsilon''(x, y, d, t) \quad (8.2)$$

where x, y, d, t are the space-time coordinates, d is depth from the land surface, m

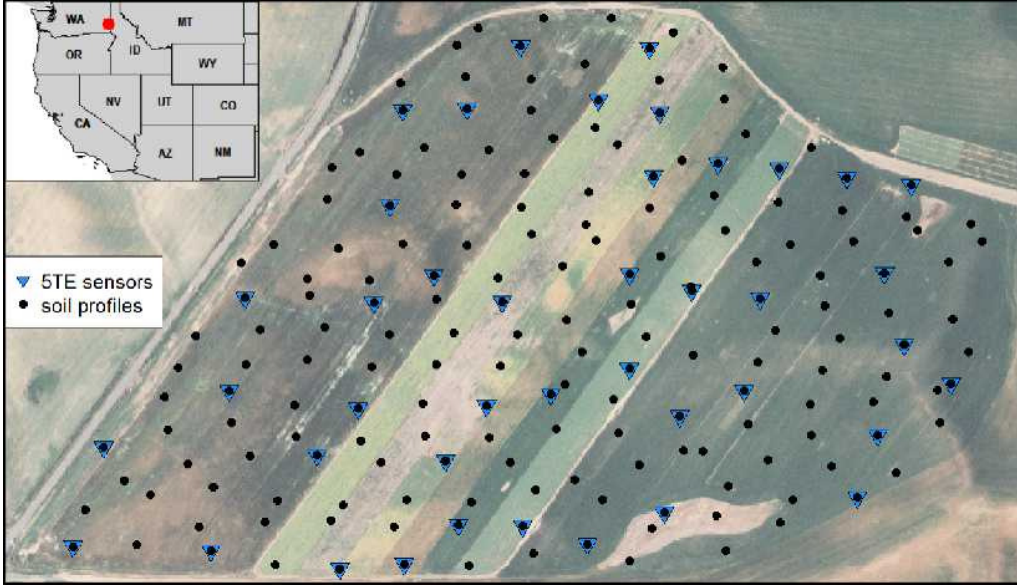


Figure 8.1: Cook Agronomy Farm overview map with soil profile sampling points (dots) and instrumented locations (triangles). A total of 210 sensors (42 locations \times 5 depths) have been collecting measurements of volumetric water content, temperature, and bulk electrical conductivity since 2009.

is the trend, $\varepsilon'(x, y, d, t)$ is the spatio-temporally correlated stochastic component and $\varepsilon''(x, y, d, t)$ is the uncorrelated noise. We model the trend (m) as a function of spatial (2D or 3D), temporal, or spatio-temporal explanatory variables (covariates, such as in Table 8.1) available over the entire spatio-temporal domain of interest.

8.2.3 3D+T random forests model

The trend model, m in Eq. 8.2, can be fitted using linear regression or some kind of Generalized Linear Model depending on the distribution of the target variable (Pinheiro and Bates, 2009). Our focus here is on fitting the trend model using random forests algorithms (Breiman, 2001a) for two main reasons. First, with random forests algorithms, the target variable does not need to assume specific distributions or adhere to linear relationships (Ahmad *et al.*, 2010; Kuhn and Johnson, 2013a). Second, random forests is advantageous for fitting a predictive model for a multivariate data set with high dimensionality. A disadvantage of random forests models, on the other hand, is that model fitting can be computationally intensive, which may become a limitation as data set complexity increases. The second disadvantage is that random forests typically tends to over-fit data sets that are particularly noisy (Statnikov *et al.*, 2008).

We model, for example changes in soil water content, in the form:

$$R> fm = VW \sim DEM + TWI + NDRE.M$$

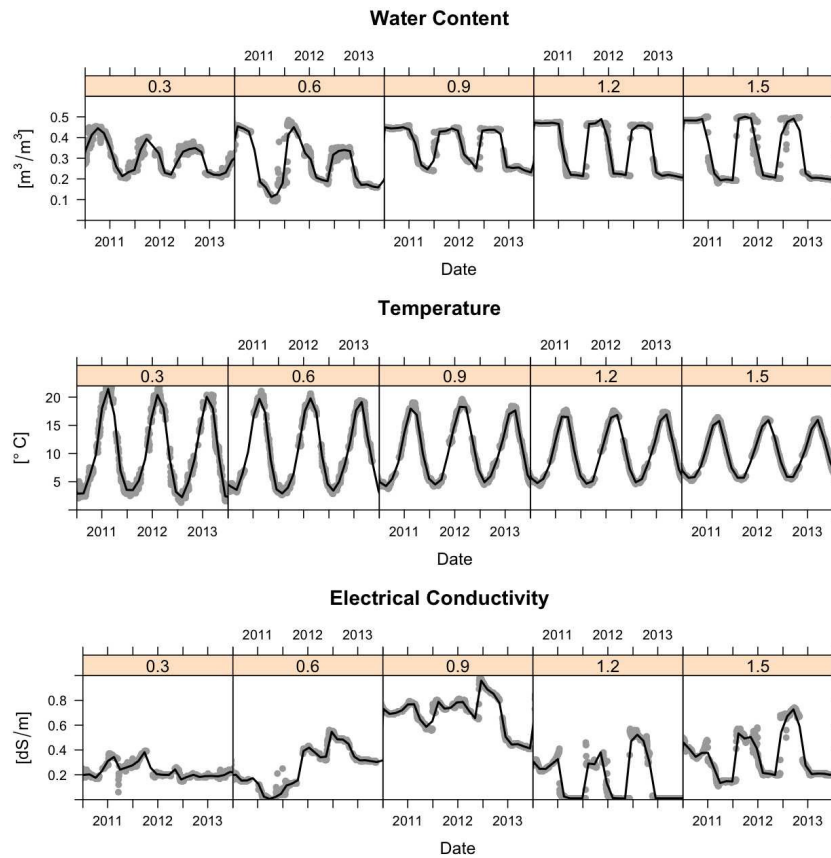


Figure 8.2: *Sensor values from five depths (0.3, 0.6, 0.9, 1.2, and 1.5 m) at one station at Cook Agronomy Farm from January 2011 — January 2014. The black line indicates locally fitted splines (here used for visualization purposes only).*

+ NDRE.Sd + Bt + BLD + PHI + Precip_cum
+ MaxT_wrcc + MinT_wrcc + cdayt + Crop

where DEM + TWI + ... + Crop are the covariates (see also Table 8.1) both measured at the same x, y, d, t locations, VW is the volumetric water content, and cdayt is the transformed cumulative day, computed as:

$$cdayt = \cos \left([t_D - \phi] \cdot \frac{2\pi}{365} \right) \quad (8.3)$$

where t_D is the linear date (cumulative days), ϕ is the time delay from the coldest day and a trigonometric function is assumed to model seasonal fluctuation of daily temperature. The predictive model, based on the spatio-temporal regression matrix (regm.VW) is:

```
R> rfm.VW <- randomForest(fm, data = regm.VW)
```

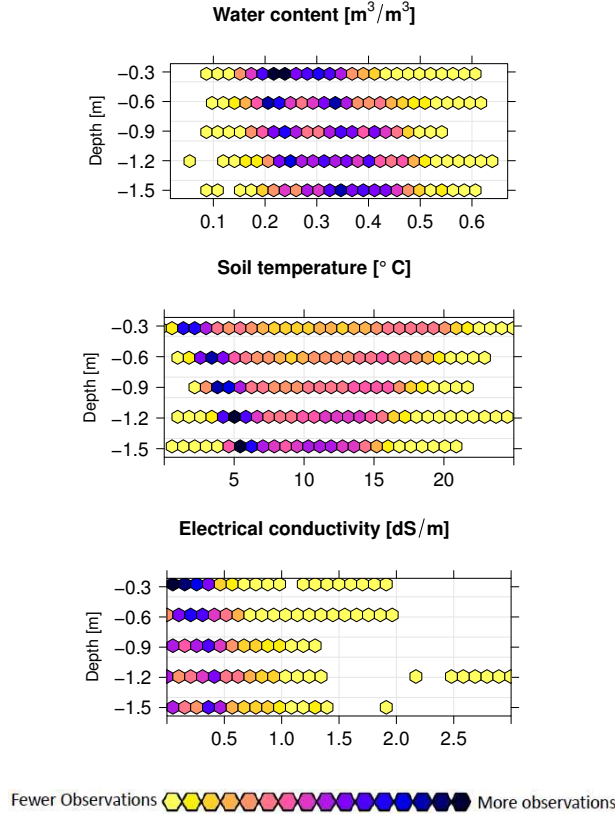


Figure 8.3: *Distribution of observations (based on all dates) for water content, temperature, and electrical conductivity across soil depth.*

The random forests prediction model from above can be used to generate predictions for any position in space and time, provided that all covariates are available at that location, but it does not provide inference on the mean trend and spatio-temporal correlation structure as in a regression-kriging model that has interpretable parameters.

In theory, 3D+T residuals of this model could be further analyzed for spatio-temporal auto-correlation and used for kriging. However, in this specific study, examination of residuals obtained from the random forests models for all three variables revealed the absence of any correlation structure over horizontal space (x, y) . Since the random forests models explained a high amount of the variability in the data ($>90\%$ for all three response variables), all residual variation was considered to be uncorrelated noise ($\varepsilon''(x, y, d, t)$ in Eq. 8.2).

8.2.4 3D+T kriging model

To explore an alternative approach to spatio-temporal random forests modeling, we developed a 3D+T regression-kriging model based on existing geostatistical methods. In this case, we use the same model as in Eq. 8.2, except we do not use any gridded or meteorological covariates to explain the trend model

Table 8.1: *Cook Agronomy Farm data set spatio-temporal covariates. DEM — Digital elevation model, TWI — SAGA wetness index, NDRE.M — Normalized Difference Red Edge Index (mean), NDRE.sd — Normalized Difference Red Edge Index (s.d.), Bt — Occurrence of Bt horizon, BLD — Bulk density of soil, PHI — Soil pH, Precip_cum — Cumulative precipitation in mm, MaxT_wrcc — Maximum measured temperature, MinT_wrcc — Minimum measured temperature, Crop — Crop type. Response variables include VW — soil volumetric water content in m³/m³, C — soil temperature in °C, and EC — soil bulk electrical conductivity in dS/m.*

	2D	depth	time	spatio-temporal	support	size
Code	(x, y)	(d)	(t)	$\Delta x, y$	Δd	Δt
DEM	✓			10 m	0 m	>10 yrs
TWI	✓			10 m	0 m	>10 yrs
NDRE.M	✓			10 m	0 m	3 yrs
NDRE.sd	✓			10 m	0 m	3 yrs
Bt	✓	✓		10 m	0.3 m	>10 yrs
BLD	✓	✓		10 m	0.3 m	>10 yrs
PHI	✓	✓		10 m	0.3 m	>10 yrs
Precip_cum			✓	spatially constant	0 m	1 d
MaxT_wrcc			✓	spatially constant	0 m	1 d
MinT_wrcc			✓	spatially constant	0 m	1 d
Crop	✓		✓	10 m	0 m	1 year
VW, C, EC	✓	✓	✓	42 points on 0.37 km	0.3 m	1 d

(m). Instead, to model the observed water content, temperature, and electrical conductivity, we only use simple seasonal detrending. Because annual patterns of weather conditions influence these soil properties in a systematic way (see Figure 8.2), detrending is necessary before we can apply any kriging. Moreover, because strength of seasonality decreases with depth and shows some delay in time, separate seasonal models were fit at each depth. Daily soil temperatures throughout the year nicely follow a sine curve with intercept c , amplitude a and shift b for the day of the year t_D^* (1 to 365) given by:

$$s_C(t_D^*) = c + a \cdot \sin\left(\frac{b + t_D^*}{365} \cdot 2 \cdot \pi\right) \quad (8.4)$$

The other two variables, water content and electrical conductivity, require a

somewhat more complex function because values are fairly stable during periods of crop inactivity. These correspond to sustained minima during the dry season (late summer to autumn) and sustained maxima after winter recharge (late winter to spring). The seasonal function for these variables is:

$$s_V(t_D^*) = c + a \cdot \cos(breaks(t_D^*) \cdot \pi) \quad (8.5)$$

with:

$$breaks(t_D^*) := \begin{cases} 1 + \frac{t_D^* + 365 - b_4}{b_1 + 365 - b_4} & , t_D^* \leq b_1 \\ 0 & , b_1 < t_D^* \leq b_2 \\ \frac{t_D^* - b_2}{b_3 - b_2} & , b_2 < t_D^* \leq b_3 \\ 1 & , b_3 < t_D^* \leq b_4 \\ 1 + \frac{t_D^* - b_4}{b_1 + 365 - b_4} & , t_D^* \leq b_4 \end{cases}$$

where $1 \leq b_1 < \dots < b_4 \leq 365$ are four consecutive break points during one year, which resemble the on- and offset of sustained minima and maxima. Hence, the function $\cos(breaks(t_D^*) \cdot \pi)$ connects two plateaus at 1 (from b_1 to b_2) and -1 (from b_3 to b_4) with smooth transitions along a stretched cosine curve. The parameters c and a in Eq. 8.5 correspond to an intercept and amplitude respectively, V indicates the variable (water content or electrical conductivity).

The models in Eq. 8.4 and Eq. 8.5 use purely mathematical functions that can be used to describe the seasonality of this data set. An alternative approach would be to use the daily mean value of sensor readings as the trend. We were interested in using these parameters to learn about how the seasonal trends of the measured soil properties change across depths. In analyses where such interpretation is unnecessary, the simpler approach may be adequate.

Assuming that the remaining residual is normally distributed and has zero mean, only its variance-covariance remains to be specified. To tackle the 3D+T data set, we assume a metric covariance model over horizontal and vertical distances after an isotropy scaling has been applied. The more general set-up would yield a 3D variogram surface in 4-dimensional space ($\gamma \sim \text{horizontal distance} + \text{depth} + \text{time}$) and can thus be reduced to the simpler 2D surface ($\gamma \sim \text{3D distance} + \text{time}$).

In order to obtain an objective estimate of the anisotropy ratio between horizontal and vertical distances, we calculated 2D empirical variograms where each day is used as a repetition of the process (i.e. distances are only calculated within each day and not across time). Based on this variogram surface, a pure metric

model can be estimated and its anisotropy scaling can then be used to construct pseudo 3D data where the depth value has been rescaled by the anisotropy ratio.

The sum-metric variogram structure for the spatial, temporal, and spatio-temporal (*‘joint’*) components, treated as mutually independent, is defined as (Heuvelink *et al.*, 2012):

$$\gamma(h, u) = \gamma_S(h) + \gamma_T(u) + \gamma_{ST}(\sqrt{h^2 + (\alpha \cdot u)^2}), \quad (8.6)$$

where $\gamma(h, u)$ is the semivariance of variable Z for 3D distances in space (h) and in time (u), γ_S, γ_T are spatial and temporal components respectively, each with a sill, range, and nugget. The joint space-time component, γ_{ST} , also includes a parameter for the conversion of temporal separation (u) to spatial distance (h), denoted α . Variogram parameters are estimated from the observations and then fit with a metric semivariance function, used in kriging to predict Z at unobserved spacetime points. For example, kriging predictions are produced from water content observations as:

```
R> svgmVW3DT <- variogramST(resid~1, VW.st)
R> fvgmVW3DT <- fit.StVariogram(svgmVW3DT,
                                vgmST("sumMetric",
                                space=vgm(sill, model, range, nugget),
                                time=vgm(sill, model, range, nugget),
                                joint=vgm(sill, model, range, nugget),
                                stAni=ratio)
R> predVW.resid <- krigeST(resid~1, VW.st, Pred.st,
                           fvgmVW3DT)
```

where `resid~1` defines the sample variogram for the water content residuals after detrending, which are stored in the `spacetime` object, `VW.st`. The sample variogram `svgmVW3DT` is used to fit a 3D+T sum-metric model, `vgmST`, wherein the variogram for each component is defined with user inputs for initial model parameters (partial sill, model type, range, nugget, and anisotropy ratio α), based on inspection of the sample variogram. The fitted variogram `fvgmVW3DT` is then used to make predictions at unobserved locations, stored in a `spacetime` object `Pred.st`. The residual predictions `predVW.resid` are added to the seasonal trend to obtain predicted water content at any spacetime point. The formulas of kriging in the spatio-temporal domain do not differ fundamentally in a mathematical or statistical sense from those of spatial kriging (Heuvelink *et al.*, 2012).

8.2.5 Cross-validation

We run cross-validation for the two spatio-temporal prediction approaches (3D+T random forests model and 3D+T kriging after detrending) separately. Moreover, we run two versions of cross-validation for the random forests model:

1. 3D+T random forests prediction (RF):
 - **RF-loc**: strict cross-validation, using leave-one-station-out iterations of model fitting and validation, and
 - **RF-rnd**: simple cross-validation, by randomly subsetting spacetime points, and using 5-fold sets of model fitting and validation,
2. 3D+T regression kriging (kriging):
 - **kriging-loc**: leave-one-station-out using the fitted variogram model, then validation.

Specific details of the cross-validation methods appear below, but first, it is important to emphasize that fundamental differences between the two modeling approaches do not allow the predictions for cross-validation to be obtained in the exact same way. In particular, the RF model is informed directly by the observations rather than a parametric model. So, if observations are removed, a new model is developed, driven by the included observations. Conversely, the kriging model quantifies the variability in the data and how it changes with distance. The inherent replication of point pairs within each lag distance buffers the resulting variogram model from the removal of an observation. These differences materialize in the cross-validation steps as follows: once a RF model has been fit with all data, the same model cannot be used on a subset of the observations (a training set) to make predictions, so a leave-one-out approach for n observations requires n training models, each unique, for n predictions. This differs from the kriging cross-validation in that the same theoretical variogram model—developed from all observations—is applied to each of the n training sets to make n predictions because automatically re-fitting the variogram for each training set would be cumbersome and is unlikely to produce considerably different variogram models.

For strict cross-validation of the RF model (**RF-loc**), 42 models were iteratively trained, each using the data of 41 stations (a ‘station’ includes all five depths and all time points, a 5-variate time series). Each model was then applied to predict on the respective withheld 5-variate time series. The results of the strict cross-validation indicate predictive performance at new, unsampled locations. For simple cross-validation of the RF model (**RF-rnd**), 10 % of observations were randomly subset from the full set of spacetime points, and subject to 5-fold

cross-validation. This less-rigorous approach provides information on predictive performance when at least some observations exist at all locations, and is useful for understanding the accuracy of interpolating missing data at an existing sample location.

To validate the kriging model (`kriging-loc`), we assumed the variogram model to be known and used the fitted model for all predictions in the cross-validation. Each of the 42 stations (including all five depths) was removed from the data set in turn. This withheld 5-variate time series was then predicted using the remaining data. For computational reasons, the prediction was limited to the closest 500 spatio-temporal neighbours (using anisotropy scalings for the 3D+T distances) from a temporal window of ± 10 days for prediction.

For each variable and each model approach, we calculated standard model performance measures: root mean square error (RMSE), mean absolute error (MAE), mean error (ME), and coefficient of determination (R^2) for observations and predictions obtained in cross-validation. As a baseline comparison, spatially constant predictions were made from the seasonal models alone (Eq. 8.4 and Eq. 8.5) for each variable and each depth. The same four cross-validation statistics were computed for these predictions. Although we do not apply exactly the same cross-validation procedures to the two methodological approaches, we assume that the cross-validation results will reveal useful information about each model's performance.

8.2.6 Software implementation

All analysis was conducted in R (R Core Team, 2014) unless otherwise noted in the text. Preparation of sensor network data and covariate data was assisted by the following packages: `aqp` (Beaudette and Roudier, 2013), `gdata` (Warnes *et al.*, 2014), `GSIF` (Hengl *et al.*, 2014b), `gstat` (Pebesma and Gräler, 2013), `plyr` (Wickham, 2014), `raster` (Hijmans *et al.*, 2014), `rgdal` (Bivand *et al.*, 2014), `RSAGA` (Brenning, 2013), and `spacetime` (Pebesma, 2012). The `randomForest` package (Liaw and Wiener, 2002) was used for the RF modeling. The kriging approach was mainly based on the `gstat` package (Pebesma, 2004) in combination with the `spacetime` package (Pebesma, 2012). The `lattice` (Sarkar, 2008) and `plotKML` (Hengl *et al.*, 2015) packages were used for data visualization.

A subset of this data set (for the period Jan. 1, 2011 — Dec. 31, 2012) and example code for the main processing steps has been added to the `GSIF` package (Hengl *et al.*, 2014b) for demonstration and can be obtained by calling `?cookfarm` after loading the package.

8.3 Results

8.3.1 3D+T random forests model

The importance plots for predicting water content, temperature, and electrical conductivity with the RF models are shown in Figure 8.4. The covariates with higher importance will influence the prediction more if randomly permuted within the model. The mean decrease in accuracy metric (%IncMSE) indicated that the cumulative date was the most important predictor for all three variables, followed by crop identity for water content and electrical conductivity, and soil pH and crop identity for soil temperature. The decrease in mean squared error (IncNodPurity) also indicated that cumulative day was important for modeling water content, and all three weather covariates were important for soil temperature. Soil properties (pH, Bt presence, and bulk density) were most important for modeling bulk electrical conductivity by the same metric.

The `randomForest` package reported that the RF models, based only on covariate data, explain 93 % of the variance in water content, 98 % in temperature, and 93 % in conductivity observations. As described in section 8.2.3, we did not fit space-time variograms to the residuals because residual variation did not display any strong spatio-temporal correlation. Further processing would produce pure-nugget variograms (of uncorrelated noise), which do not impart any additional explanatory power.

Prediction surfaces for water content (for the first day of five months in 2012) produced directly from the RF model are shown in Figure 8.5. This period of time represents the growing season, when large changes in water content occur as crops develop and rapidly extract soil water. Prediction maps for water content, soil temperature, and electrical conductivity for the whole period of observation (spacetime prediction stacks) can be obtained by contacting the authors.

8.3.2 3D+T kriging model

Fitted parameters for the seasonality functions (Eq. 8.4 and Eq. 8.5) are listed in Table 8.2. The seasonal effects varied by depth: as depth increases, the change in soil properties was delayed, and the amplitude of the change, on average, increased for water content and decreased for soil temperature. For electrical conductivity, the amplitude was highest at 0.9 m. High temperatures corresponded with low water content and associated conductivity.

Table 8.3 lists the h/v ratios for horizontal-vertical distance scaling, as well as the variogram parameters for each variable. We set the h/v ratios so that 1 m in depth horizontally corresponded to 21 m for water content, 516 m for soil

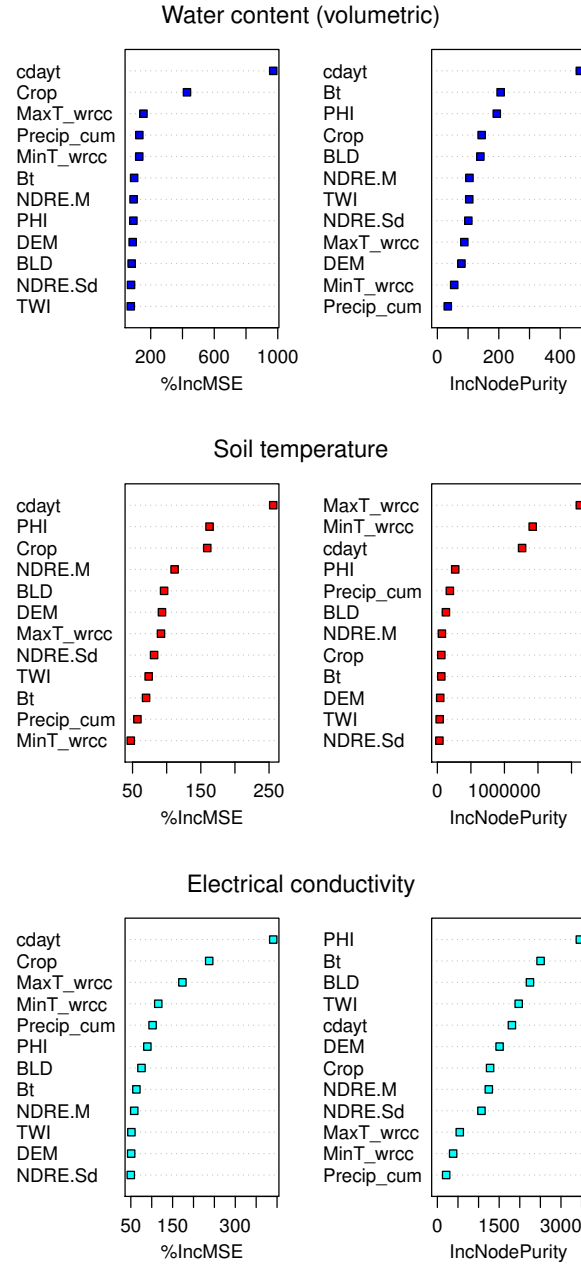


Figure 8.4: Importance plots (covariates sorted by importance) derived using the *randomForest* package (Liaw and Wiener, 2002). %IncMSE is the mean decrease in accuracy; IncNodePurity is the decrease in mean squared error.

temperature, and 53 m for electrical conductivity.

The water content and electrical conductivity variogram models only contained the metric component (γ_{st}), each with four parameters (sill, range, nugget, and the anisotropy parameter α), while soil temperature used a sum-metric model with spatial, temporal, and joint components as in Eq.(8.6). The lack of pure spatial and temporal components in water content and conductivity indicated that these correlation structures appeared to be sufficiently modeled through a

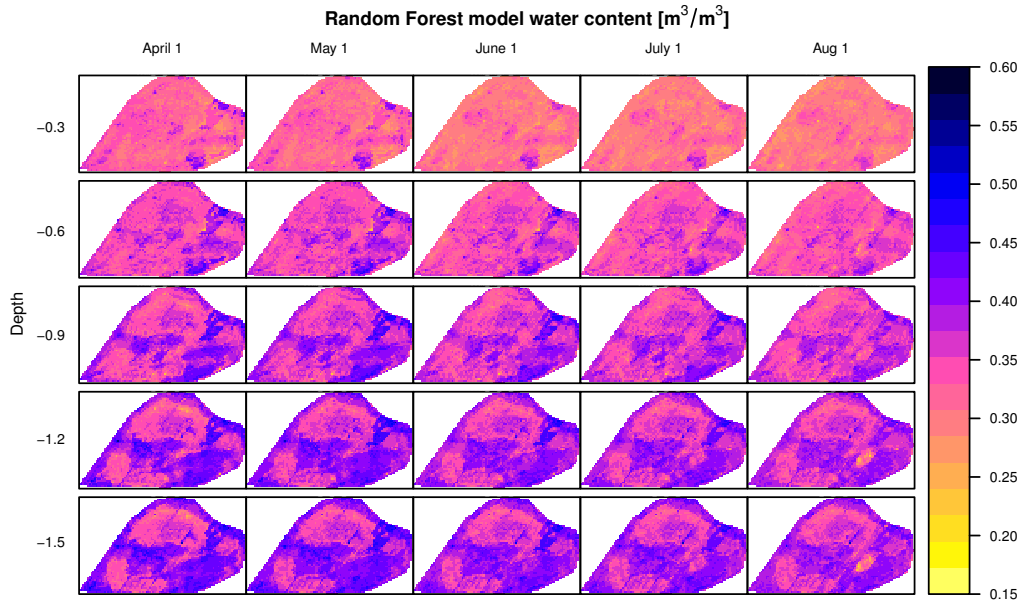


Figure 8.5: *Spatio-temporal predictions of soil water content at Cook Agronomy Farm for the growing season in 2012 using the random forests (RF) model. Note that relative changes in water content are accurate, but absolute sensor readings require correction.*

metric model. In all three cases, the correlation in time was stronger over larger separation distances, indicated by anisotropy ratios (sp/t) that were less than one. For example, correlation at 1 m was equal to correlation at 5 days for water content, 2 days for temperature, and 17 days for electrical conductivity. This translates to the inclusion of more temporal neighbors than spatial neighbors when making kriging predictions. Sample and 3D+T fitted variograms are depicted in Figure 8.6, along with isolated 3D spatial and temporal components. Please note that the optimization of the spatial and temporal components of each 3D+T variogram is done based on the full spatio-temporal model. Hence, the fit represents the entire variogram surface. As a result, the individual space and time components may not intersect the sample data and appear as a poor fit compared with the overall surface. Prediction surfaces for water content during the 2012 growing season were also created from the 3D+T kriging model, shown in Figure 8.7.

8.3.3 Model accuracy

For all three variables, Figure 8.8 shows hexbin plots of observed versus predicted values with the full RF model, strict cross-validation of the RF model (RF-loc), and cross-validation of the kriging model (kriging-loc). Table 8.4 lists the global cross-validation statistics for the two models in addition to the

Table 8.2: Parameters of the seasonality functions (Eqs. 8.4 and 8.5) for water content (VW), soil temperature (C) and electrical conductivity (EC) at each depth. The parameters represent the intercept (c), amplitude (a) and shift (b) in seasonal effects at each depth.

var.	depth	c	a	b	b_1	b_2	b_3	b_4
VW	0.3 m	0.26	0.06		45	128	223	257
	0.6 m	0.29	0.06		66	153	228	280
	0.9 m	0.31	0.06		71	152	258	282
	1.2 m	0.32	0.05		71	169	266	289
	1.5 m	0.35	0.03		117	186	205	245
C	0.3 m	9.7	-8.9	63				
	0.6 m	9.6	-7.4	55				
	0.9 m	9.5	-6.3	46				
	1.2 m	9.5	-5.4	38				
	1.5 m	9.4	-4.6	30				
EC	0.3 m	0.20	0.05		55	133	219	241
	0.6 m	0.31	0.08		93	160	225	248
	0.9 m	0.37	0.11		77	148	254	290
	1.2 m	0.41	0.09		84	184	225	274
	1.5 m	0.44	0.07		62	110	281	332

Table 8.3: Variogram parameters for each variable. VW is water content, C is soil temperature, EC is electrical conductivity, $\frac{h}{v}$ is the anisotropy ratio for horizontal-vertical distances (m); st-vgm is the sum-metric component of the spatio-temporal variogram; $\frac{sp}{t}$ is the anisotropy ratio between spatial and temporal (m/days) distances (α); sill, range, and nugget are variogram parameters; and the semivariance function of each model is either Exponential (Exp) or Spherical (Sph). Sill and nugget units are the same as the measured variable.

var.	$\frac{h}{v}$	st-vgm	$\alpha = \frac{sp}{t}$	sill	model	range	nugget
VW	21	joint	0.20	0.005	Exp	32 m	0
C	516	space		0.26	Exp	97 m	0.39
		time		4.69	Exp	147 days	0
		joint	0.48	0.27	Sph	20 m	0
EC	53	joint	0.06	0.06	Exp	21 m	0

spatially constant seasonal models used for detrending.

The goodness of fit between observations and predictions using the full RF model was >90 % for all three variables. However, under strict cross-validation

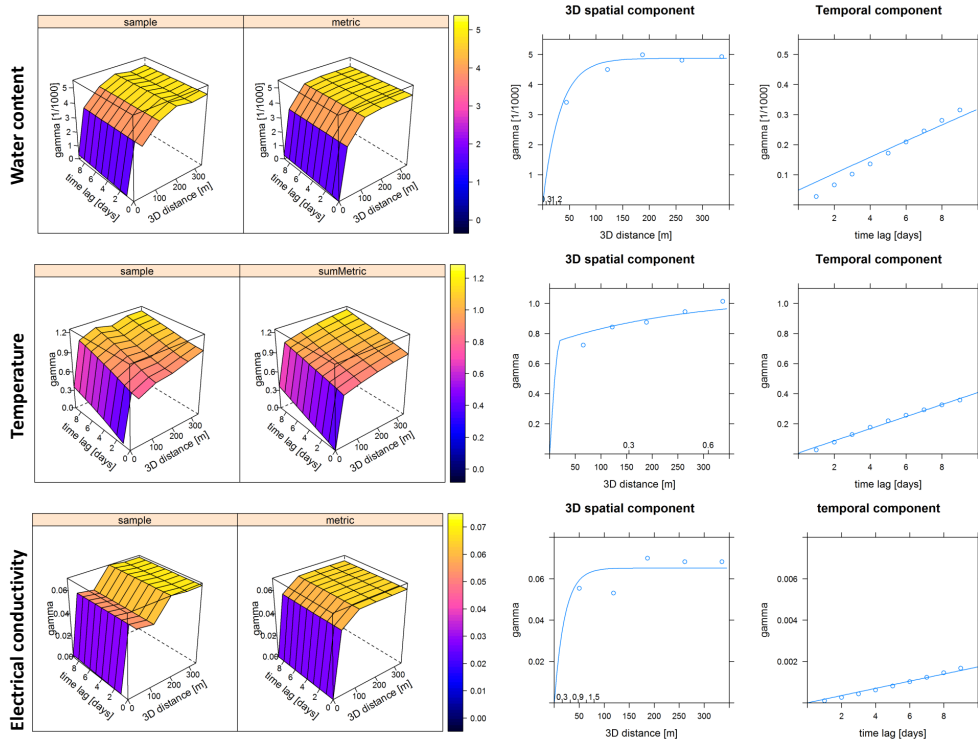


Figure 8.6: *Spatio-temporal sample variogram, metric variogram, and isolated 3D spatial and temporal components for water content, temperature, and electrical conductivity. The double axis on the 3D variogram illustrates the relationship between vertical and horizontal depths.*

(RF-loc), the predictive power of the RF model decreased, especially for water content (34%) and conductivity (5%). The R^2 values for soil temperature remained high in cross-validation. The less rigorous cross-validation procedure (RF-rnd) demonstrated stronger predictive power and lower error for all three variables, with 86%, 97%, and 88% of variability explained for water content, temperature, and conductivity, respectively.

The seasonal models alone predicted all variables well, with the kriging models only capturing a bit more variability. As with the RF model, the kriging model was most successful at predicting soil temperature. The R^2 of the kriging model for the highly variable electrical conductivity was low at 18%. Both the RF and kriging models had difficulty predicting the infrequently high conductivity values.

8.4 Discussion

8.4.1 Model performance

In this paper we examined two approaches to producing continuous predictions from 3D+T point observations of three dynamic soil variables, measured

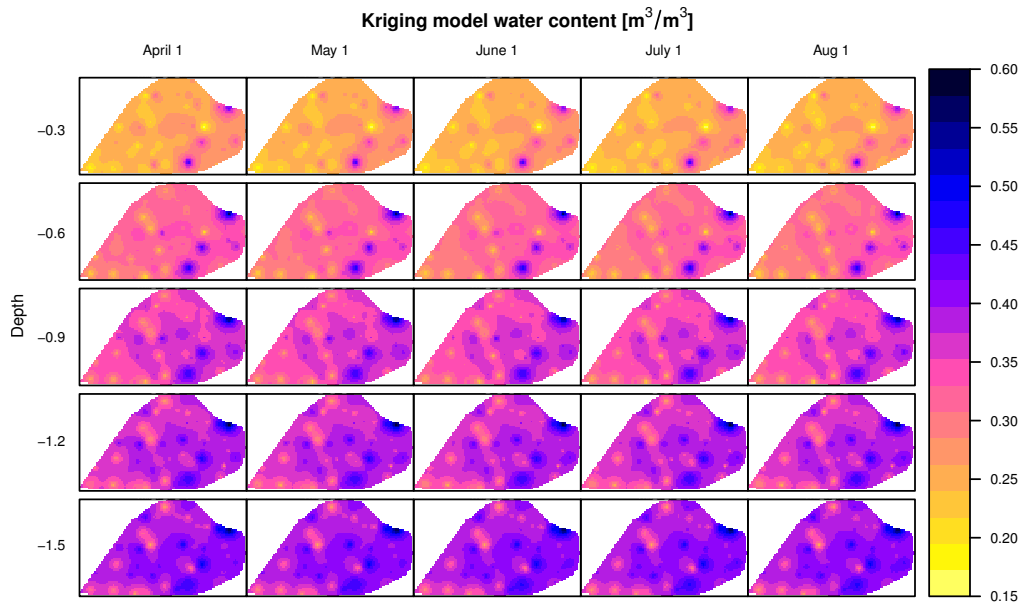


Figure 8.7: *Spatio-temporal predictions of soil water content at Cook Agronomy Farm for the growing season in 2012 using the kriging model. Note that relative changes in water content are accurate, but absolute sensor readings require correction.*

daily at the field scale, by a 3D sensor network, for multiple years, and on complex terrain that hosts rotating cropping systems. First, we assembled a highly dimensional spatio-temporal regression matrix, and when fit with random forests algorithm, covariates successfully explained the variability in observations. All of the measured variables displayed seasonal patterns (Figure 8.2), so temporal covariates explained much of the variability in the observations. Cumulative day was an important covariate for all three soil variables, as was crop identity. At Cook Agronomy Farm, the field is divided into multiple strips, which are the basis for crop rotations. Different cropping systems have different patterns of water use, biomass production, rooting depth, and influences on the soil surface e.g. shading, residue production, and interception of precipitation (Al-Mulla *et al.*, 2009; Qiu *et al.*, 2011). These characteristics are likely responsible for the differences in dynamic soil properties between the strips, and from year to year—thus, they can explain both spatial and temporal variability.

We expected precipitation to be an important predictor of soil water content; however, weather covariates, were only deemed important according to the decrease in mean squared error metric. While precipitation is the only source of soil water in this dryland agricultural system, evapotranspiration also plays an important role in controlling soil water content, along with terrain and soil properties (Cantón *et al.*, 2004; Hébrard *et al.*, 2006). Perhaps inclusion of estimated

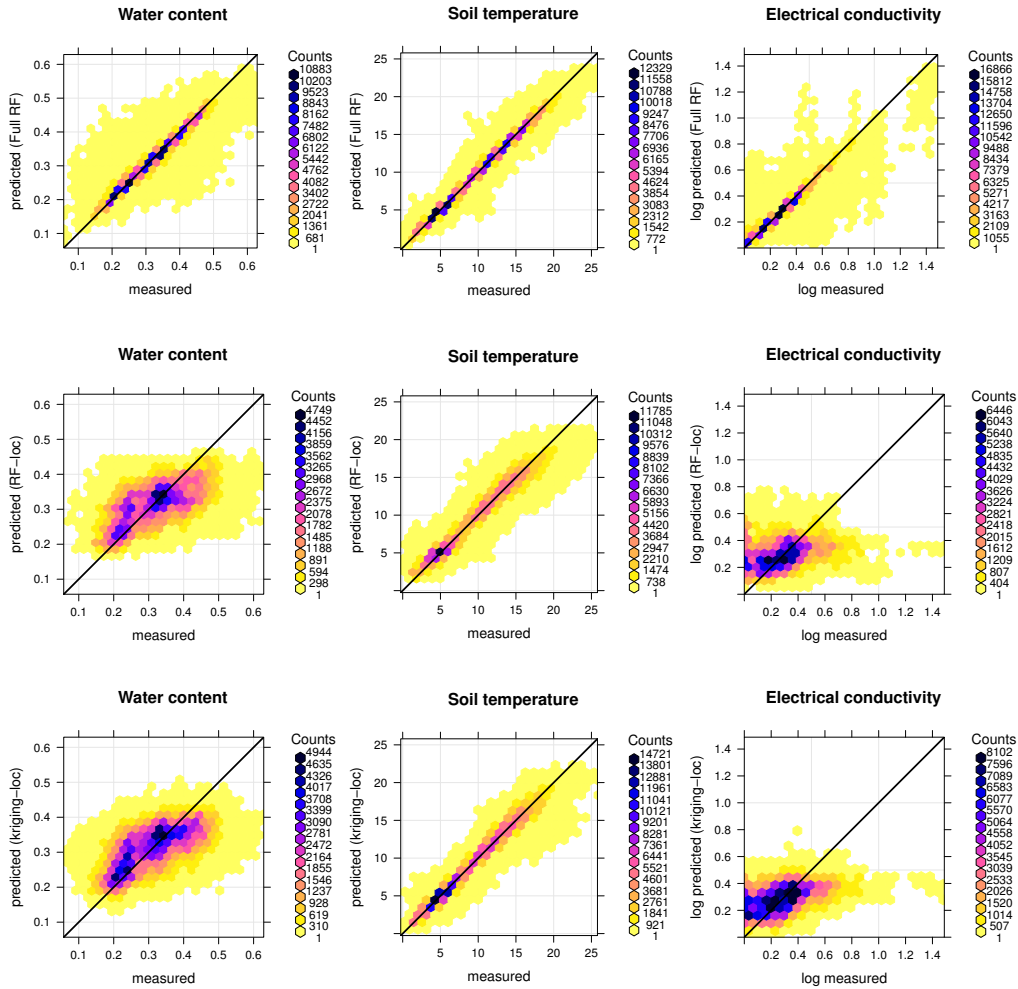


Figure 8.8: Hexbin plots for observed and predicted values for the full RF model showing goodness of fit (top), strict cross-validation of the RF model (center), and of the kriging model (bottom).

evapotranspiration as a covariate, as in Jost *et al.* (2005), would complement our covariate set in predicting soil water. The confounding and interacting effects of weather, terrain, and soil properties that influence soil water content were likely not recognized by the random forests model, as covariates are assessed individually.

Similarly, we expected air temperatures to be important in explaining variability in soil temperature. Daily minimum and daily maximum temperatures indeed had high importance, according to one of the rankings; however, air temperatures may not be representative of heat fluxes at the soil surface, due to crop influences mentioned above.

Soil bulk electrical conductivity is correlated with soil moisture, organic matter, soil salinity, and soil texture (Friedman, 2005). Accordingly, we expected covariates that are important in predicting soil water content to also predict con-

Table 8.4: Global cross-validation statistics including the spatially constant predictions based on the fitted seasonality functions, the kriging model (kriging-loc), and two sets of statistics for the RF model (RF-loc and RF-rnd). *VW* is water content, *C* is soil temperature, and *EC* is electrical conductivity, *RMSE* is root mean squared error, *MAE* is mean absolute error, *ME* is mean error, and R^2 is coefficient of determination. The R^2 for *EC* was calculated on the log scale, due to a skewed distribution.

var.	approach	RMSE	MAE	ME	R^2
VW	season	0.08	0.06	0.00	0.31
	kriging-loc	0.07	0.06	0.00	0.37
	RF-loc	0.07	0.06	0.00	0.34
	RF-rnd	0.03	0.02	0.00	0.86
C	season	1.37	1.03	0.00	0.93
	kriging-loc	0.98	0.70	0.01	0.96
	RF-loc	1.30	0.96	0.06	0.93
	RF-rnd	0.94	0.67	0.00	0.97
EC	season	0.27	0.20	0.00	0.13
	kriging-loc	0.27	0.19	-0.01	0.18
	RF-loc	0.31	0.21	0.00	0.05
	RF-rnd	0.10	0.05	0.00	0.88

ductivity, in addition to soil properties related to soil texture (bulk density and Bt horizon presence). These covariates were ranked with high importance in the RF model.

According to the strict cross-validation, the predictive success of the RF model decreased as the variability of the target variable increased. This suggests that the model was sensitive to the micro-scale variation in the data, rather than capturing the general spatio-temporal trend of the data. While the random forests algorithm generally tries to resist overfitting (Breiman, 2001a), instances of overfitting have been documented (Statnikov *et al.*, 2008). Conversely, under the simple cross-validation, the predictive power was strong. Clearly, the inclusion of at least some spacetime points at a location were crucial for making predictions at each location using the random forests algorithm. The 42 instrumented stations are intentionally stratified across the terrain and soil feature space, and no two locations are the same. We suspect that the stations are sparse enough across the complex landscape of Cook Agronomy Farm that predicting new, unique locations occurs with higher error. It would be interesting to see if additional sensor stations would improve predictive power, and/or if model performance was improved in a more uniform study area. Identifying the optimal sample size for high predictive accuracy in a complex study area is a question that still needs to be

addressed. Through this analysis, we have also realized that there are multiple ways of dividing the data set for cross-validation of these models — each providing different information about dependence across space, time, or both. Here, we applied validation methods familiar to spatial analysis, but we suspect that these methods are limited for handling complex 3D+T data. In the future, we hope to explore cross-validation methods that better assess predictive power through space, time, and their interaction.

We also expanded the kriging framework to accommodate the 3D+T data. These models first required that we de-trend the data with depth-dependent seasonality functions. The parameters of the seasonality functions that we fit demonstrate that all three variables experienced a temporal delay as soil depth increases. These results reflect the infiltration process during soil water recharge, and later in the season, water draw-down by crop roots at increasing depths. Similarly, seasonal soil temperature changes experienced a lag as soil insulation increases with depth. Soil electrical conductivity followed a similar seasonal pattern as water content, but with the largest minima and maxima at depths where clay horizons occur. These depth-dependent temporal patterns explained most of the variability in all three variables, akin to in the RF model.

3D+T variograms parameters indicated that spatial heterogeneity was high, while temporal correlation was stronger over longer separation distances (spatial range parameters were shorter than temporal range parameters). Soil temperature was correlated over shorter time periods, but was more constant over vertical space (as indicated by the h/v ratio). Water content was correlated over longer time periods, but over shorter vertical space. This translates to temperature changes in the soil occurring at a faster rate than changes in water content, but water content was more variable across 3D space. Electrical conductivity was the least dynamic of all, because it is partially dependent on static soil properties e.g. clay content (Corwin and Lesch, 2005). The presented 3D+T kriging approach only uses *day of the year* as a covariate. Including some of the many covariates used in the random forests approach to define the regression trend might also improve the performance of regression-kriging for this data set.

For both modeling approaches, temporal patterns explained most of the variability in the observations, while spatial components were secondary. Spatial heterogeneity is high at Cook Agronomy Farm, with hilly terrain, variable soil horization, and multiple crop rotations. Our ability to predict this spatial complexity with high precision was limited with only 42 stations. Thus, the high temporal sampling density within this data set seems to be more important to our modeling efforts.

8.4.2 Interpretation of model predictions

All three soil variables show interesting patterns through the soil profile, across horizontal space and time. The range of water content was higher in the shallower soils, which are exposed to extremely wet and extremely dry conditions. Additionally, on average, soil water was retained in deeper soil, relative to shallower depths. This was similar to soil temperature, where deeper soils are insulated from extreme air temperatures, in both cold and warm seasons. Electrical conductivity was variable through the profile, with some higher values occurring in shallow soils, possibly due to fertilizer application (De Neve *et al.*, 2000; Eigenberg *et al.*, 2002). High values also occurred at the 1.2 m depth, which may be an indication of accumulated carbonates or other materials. It is important to note that the electrical conductivity readings represent the conductivity of the bulk soil (including solid and liquid states). These values may be converted to conductivity of the soil solution, which would be of interest for assessing soil salinity specifically related to land and vegetation management. Soil solution salinity is calculated from the bulk conductivity using the dielectric permittivity, soil temperature, and water content measured by the sensors (Decagon Devices, Inc., 2014; Hilhorst, 2000). Depending on the research question, either bulk or soil solution conductivity could be interpolated with the methods described here. It is possible that soil solution electrical conductivity may display less variability and be easier to predict in space and time.

The prediction surfaces produced from the RF model showed more fine-scale variability, compared to the kriging predictions. This was a result of the inclusion of crop, terrain, and soil covariates in the predictive model. Within the prediction surfaces, spatial patterns of covariate features are visible; particularly for the covariates that ranked with high importance in the models (e.g. cropping strips and Bt presence in Figure 8.5). The only spatial information provided by the kriging model was the spatio-temporal correlation around each sample point—causing the speckled appearance of the map. Nevertheless, in both cases, we can see that deeper soil retained water when shallow soil was dry late in the growing season. The seasonality and draw-down of soil water was more apparent in the RF model predictions, than in the kriging predictions, particularly in deeper soil. Certainly, the kriging predictions provide a more spatio-temporally smoothed representation of the response variables, compared with the RF model.

8.4.3 Final conclusions and future directions

We have demonstrated two approaches for interpolating dynamic 3D+T soil data. We observed that both models were highly successful in predicting soil temperature and that the predictive power decreased as property variability increased

— particularly when data from a station was entirely absent. The temporal components in each model contributed most to explaining all three soil variables across depth, emphasizing the importance of the seasonal changes in this data set. Modeling changes in soil properties through time is, perhaps most interesting for variables where such change can be observed at temporal scales of a few days to a few years (Figure 8.9). Certainly, dynamic properties that irregularly or erratically change will require innovative modeling approaches for explaining such temporal behavior.

It should be noted that these methods are experimental and invite modification and improvement. The results presented here are specific to the Cook Agronomy Farm data set; the work serves as a case study for exploring 3D+T interpolation approaches, and a basis upon which we can build. We observed that temporal autocorrelation and time (day of the year) largely contribute to the portion of variation that we can explain. A future direction could include combining a random forests model with residual kriging. Given such a large data set, we can experiment with thinning the regression matrix to remove spatial and/or temporal correlation from the random forests model, and integrate those predictions with spatio-temporal kriging.

Development of 3D+T models to create continuous predictions from point data will allow dynamic soil properties to be incorporated into spatially-explicit process and biophysical models. These spatio-temporal predictions of soil water content, temperature, and electrical conductivity, as well as the 3D maps of basic soil properties such as pH and bulk density, can inform precision agricultural practices. All these soil variables can assist in understanding site specific characteristics of Cook Agronomy Farm, such as crop performance, or risk of fertilizer loss to the groundwater or the atmosphere. The fitted initial spatio-temporal models can also be used to optimize soil monitoring networks (Heuvelink *et al.*, 2012) and/or recommend sampling and modeling strategies for properties that might co-vary through space and time.

3D+T predictions of key soil properties also assist in visualizing dynamic below-ground properties, which, unlike above-ground properties, cannot be observed with photography or remote sensing. Time-lapse animations of 3D soil properties provide information that is difficult to access through static, piecewise, representations. As a supplement to this paper, we have included KML (Keyhole Markup Language) files to illustrate how 3D+T predictions can be visualized in an interactive browser such as Google Earth.

Modeling data in 3D+T is not limited to soil or agricultural applications. Any point data collected in 3D and through time could benefit from 3D+T interpolations. In short, 3D+T models allow us to visualize and access knowledge about dynamic properties that are difficult to directly observe. As technologies for mon-

itoring ecosystem properties improve and high resolution spatial data collection becomes cheaper and easier, the majority of soil maps could become 3D+T.

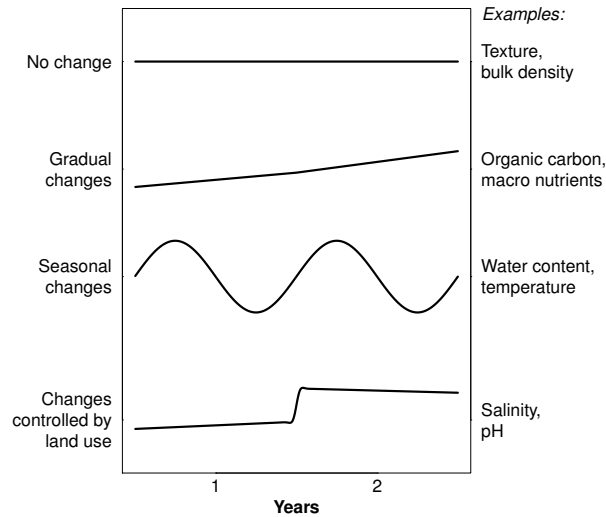


Figure 8.9: *Types of soil variables in terms of temporal stability or change.*

Acknowledgments The authors wish to thank David Huggins, Dave Uberuaga, Erin Brooks, Colin Campbell, Doug Cobos, Maninder Chahal, and Matteo Poggio for developing and maintaining the sensor network and collecting covariate data at Cook Agronomy Farm. This project was funded by the Site-Specific Climate Friendly-Farming project, provided by USDA-NIFA award #2011-67003-30341. This work was possible thanks to the software packages for organizing, visualizing, and analyzing soil data (Beaudette and Roudier, 2013) and spatio-temporal data (Pebesma and Bivand, 2013; Pebesma, 2012; Pebesma and Gräler, 2013). We are grateful to the R open source software community (R Core Team, 2014) for providing and maintaining numerous spatial and spatio-temporal analysis packages used in this work.

Chapter 9

Soil respiration and its temperature sensitivity: rapid acquisition using mid-infrared spectroscopy

Nele Meyer (1), Hanna Meyer (2), Gerhard Welp (1), Wulf Amelung (1)

(1) Institute of Crop Science and Resource Conservation (INRES), Soil Science and Soil Ecology, University of Bonn, 53115 Bonn, Germany

(2) Environmental Informatics, Faculty of Geography, Philipps-University Marburg, 35037 Marburg, Germany

Submitted to

Geoderma

Submitted 13 July 2017

9 Soil respiration and its temperature sensitivity: rapid acquisition using mid-infrared spectroscopy

Nele Meyer, Hanna Meyer, Gerhard Welp, Wulf Amelung

Abstract

Spatial patterns of soil respiration (SR) and its sensitivity to temperature (Q10) are one of the key uncertainties in climate change research but since their assessment is very time-consuming, large data sets can still not be provided. Here, we investigated the potential of mid-infrared spectroscopy (MIRS) to predict SR and Q10 values for 124 soil samples of diverse land use types taken from a 2868 km^2 catchment (Rur catchment, Germany/Belgium/Netherlands). Soil respiration at standardized temperature (25°C) and soil moisture (45% of maximum water holding capacity, WHC) was successfully predicted by MIRS coupled with partial least square regression (PLSR, $R^2 = 0.83$). Also the Q10 value was predictable by MIRS-PLSR for a grassland submodel ($R^2 = 0.75$) and a cropland submodel ($R^2 = 0.72$) but not for forested sites ($R^2 = 0.03$). In order to provide soil respiration estimates for arbitrary conditions of temperature and soil moisture, more flexible models are required that can handle nonlinear and interacting relations. Therefore, we applied a random forest model, which includes the MIRS spectra, temperature, soil moisture, and land use as predictor variables. We could show that SR can be simultaneously predicted for any temperature (5-25°C) and soil moisture level (30-75% of WHC), indicated by a high R^2 of 0.73. We conclude that the combination of MIRS with sophisticated statistical prediction tools allows for a novel, rapid acquisition of SR and Q10 values across landscapes and thus to fill an important data gap in the validation of large scale carbon modeling.

keywords Heterotrophic soil respiration; Environmental soil classes; PLSR; Random forest

9.1 Introduction

Heterotrophic soil respiration (SR) represents the second largest carbon (C) flux in terrestrial ecosystems (Raich and Schlesinger, 1992) and its accurate modeling is therefore of immense importance for reliable assessments of net ecosystem exchange. Heterotrophic soil respiration is, amongst other factors, mainly a function of soil organic carbon (SOC) quantity (Wang *et al.*, 2003), SOC degradability (Conant *et al.*, 2008), soil temperature (Kirschbaum, 1995), and soil moisture (Orchard and Cook, 1983). Due to the large number of determining factors, there are still many uncertainties in the accurate prediction of SR across various soils and weather conditions. Especially the effect of temperature on SR is one of the key uncertainties in climate change research (Kirschbaum, 2006). The temperature sensitivity of SR is often expressed as the Q10 value, which is the increase of SR by a 10°C increase in temperature (Kirschbaum, 1995; Van't Hoff, 1898). Although the Q10 value is implemented as a fixed constant in modeling approaches (e.g., 1.5 in CLM, Foereid *et al.* (2014), or 2 in CASA and TEM, Potter *et al.* (1993); Raich *et al.* (1991)), measurements have demonstrated that the temperature sensitivity is variable, with Q10 values ranging from 1 to higher than 12 (Hamdi *et al.*, 2013; Meyer *et al.*, under review). With increasing demand of implementing variable Q10 values into carbon models instead of a fixed value (e.g., Lefèvre *et al.*, 2014) the need to understand and predict spatiotemporal patterns of Q10 becomes critically important. Based on Q10 values and SR at a reference temperature, SR at any other temperature is usually calculated according to equation 9.1:

$$SR_T = SR_{REF} * Q10^{(T-T_{REF})/10} \quad (9.1)$$

where SR_T is soil respiration at the requested temperature, SR_{REF} is soil respiration at a reference temperature, T_{REF} is the reference temperature and T is the requested temperature. Consequently, both, an inaccurate Q10 value and an inaccurate SR_{REF} can lead to large over- or underestimation of carbon fluxes (e.g., Meyer *et al.*, under review; Zhou *et al.*, 2009). Hence, improvements of SR_{REF} and Q10 estimates are required for accurate predictions of carbon fluxes. Due to the time-consuming analyses, efforts to unravel the spatiotemporal variability of SR and Q10 values are often limited to small sample sizes (e.g., Reichstein *et al.*, 2003; Zheng *et al.*, 2009). An option could be the use of pedotransfer functions. However, due to the complexity of determining factors these are frequently not available or again time-consuming to develop (e.g., Fierer *et al.*, 2006). Hence, a method, which allows for a rapid and reliable estimate of SR and Q10 across the landscape, would be highly desirable. Mid-infrared spectroscopy (MIRS) coupled with partial least square regression (PLSR) has

proven its suitability for the rapid assessment of a broad range of soil properties like the content of SOC (Bornemann *et al.*, 2010; Cobo *et al.*, 2010; Janik and Skjemstad, 1995), nitrogen (Janik and Skjemstad, 1995; Zimmermann *et al.*, 2007), pH values (Cobo *et al.*, 2010; D'Acqui *et al.*, 2010; Janik and Skjemstad, 1995), texture (Cobo *et al.*, 2010; Zimmermann *et al.*, 2007), or specific carbon fractions (Bornemann *et al.*, 2008, 2010; Ludwig *et al.*, 2008). Inasmuch as SR and Q10 depend on SOC quantity and degradability, it seemed thus reasonable to speculate that it should also be possible to assess SR and Q10 values using MIRS. Earlier studies already indicated the potential of near infrared spectroscopy as a prediction method for SR (e.g., Chang *et al.*, 2001; Ludwig *et al.*, 2002; Palmborg and Nordgren, 1993). Yet, the potential of the wider range of infrared signals using MIRS to predict SR received far less attention. Further, previous spectroscopic approaches did not focus on the temporal variability of SR but predicted SR rates at a single level of soil moisture and temperature. In this study, we investigated the potential of MIRS for the high throughput estimate of SR across the landscape and for various weather conditions. First, we aimed at investigating whether a MIRS-based prediction of SR can potentially be achieved for standardized temperature and soil moisture conditions. To achieve a scaling of SR across various temperatures, we further aimed at investigating whether Q10 values are predictable by MIRS. We hypothesized that a prediction of both SR and Q10 is feasible using standard multivariate statistical procedures, e.g., partial least square regression (PLSR). To directly estimate SR beyond standardized weather conditions, in a second step, we aimed at building up a prediction model that allows for the simultaneous prediction of SR at any given temperature and soil moisture level. As the number of potential predictor variables and interactions increased, more complex statistical models were required. We hypothesized that the simultaneous prediction can be accomplished by using Random Forest modeling that is able to deal with nonlinearity and various types of predictor variables. The accomplished prediction model should represent a basis for a rapid estimation of SR across the landscape.

9.2 Material and Methods

9.2.1 Study area

The Rur River catchment is situated predominantly in western Germany and partly in the Netherlands and Belgium. The study site covers a total area of 2868 km² and exhibits distinct gradients of elevation, temperature, and precipitation. The northern part is characterized by a fairly flat area (< 100 m above sea level) with a maximum average temperature of 10.9°C and a minimum annual precipitation of 703 mm. Elevation increases towards the south where low

mountain ranges occur (Eifel, up to 700 m above sea level). Average temperature decreases to a minimum of 6.1°C and annual precipitation increases to a maximum of 1358 mm in the southeastern parts (Hijmans *et al.*, 2005; Simmer *et al.*, 2015). Intensive agriculture and urban areas dominate the flat northern parts while the southern part is mainly characterized by forests and grasslands. The Rur Catchment is the main research area of the Collaborative Research Center SFB/TR 32 (Simmer *et al.*, 2015; Vereecken *et al.*, 2016) and is also part of the Terrestrial Environmental Observatories program, TERENO (Zacharias *et al.*, 2011).

9.2.2 Soil sampling

To capture a broad range of soil properties within the catchment, we divided the catchment into environmental soil classes (ESC), which we defined as a unique combination of the factors land use, aggregated soil group, and texture (Table 9.1). For further information on this classification approach, see Meyer *et al.* (under review). Sampling was conducted during two field campaigns in two subsequent years (2015, 2016). In the first sampling campaign, we took 9 soil samples from each of the 12 most frequent ESCs. To account for possible influences of the climatic and altitudinal gradient and to avoid spatial autocorrelation, 3 of the 9 samples were taken from different locations in the northern part of the catchment, 3 from the central part, and 3 from the southern part. For a detailed description of the soil sampling design see Meyer *et al.* (under review). We complemented this sampling set by taking additional samples in the same season but one year after the first sampling campaign. This was necessary because the number of cropland and grassland samples from the first sampling campaign was comparatively low. The final sampling set comprised 124 samples including 30 cropland soils, 31 grassland soils, and 63 forest soils (including deciduous and coniferous forests) from various soil groups and texture classes and from a depth of 0-29 cm, each (Table 9.1).

All samples were sieved at field-moist conditions to 2 mm. Parts were stored at -18°C for respiration analyses, and parts were dried at 40°C for chemical analyses and MIRS measurements.

We are aware that sieving has been criticized for altering SR compared to real-world conditions (Herbst *et al.*, 2016). However, Černohlávková *et al.* (2009) and Thomson *et al.* (2010) found that the effect of sieving on SR may be neglected for field-moist samples, i.e., they recommend sieving field-moist samples, and so we did. Further, in a preliminary study, we investigated the extent by which SR and Q10 is affected by sieving (field-moist) and freeze-storage and found no significant difference (data not shown). Hence, we are confident that our results are transferrable to real world conditions.

Table 9.1: *Environmental soil classes (ESC) and sample set*

Land use	Land use sub-type	Aggregated soil group	Texture	Number of samples
Cropland	-	Terrestrial	Silt	10
Cropland	-	Terrestrial	Sand	3
Cropland	-	Stagnic	Silt	9
Cropland	-	Terrestrial	Loam	3
Cropland	-	Semi-terrestrial	Silt	5
Grassland	-	Terrestrial	Silt	11
Grassland	-	Terrestrial	Sand	2
Grassland	-	Terrestrial	clay	1
Grassland	-	Stagnic	Silt	9
Grassland	-	Semi-terrestrial	Silt	9
Forest	Deciduous	Terrestrial	Silt	9
Forest	Deciduous	Terrestrial	Sand	9
Forest	Deciduous	Stagnic	Silt	9
Forest	Deciduous	Semi-terrestrial	Silt	9
Forest	Coniferous	Terrestrial	Silt	9
Forest	Coniferous	Terrestrial	Sand	9
Forest	Coniferous	Stagnic	Silt	9

9.2.3 Determination of physicochemical soil properties

The total C and N contents of the sieved and milled soils were determined by elemental analysis (ISO 10694, 1995). All samples were free of inorganic C. Thus, total C equaled SOC. Maximum water holding capacity (WHC) and actual soil moisture were determined on field moist samples, the procedure mainly following Alef and Nannipieri (1995).

9.2.4 Soil respiration measurements and determination of Q₁₀

Samples were allowed to thaw for three days at 4°C. Afterwards, subsamples were rewetted to 30%, 45%, 60%, and 75% of water holding capacity (WHC) by adding deionized water which was homogenized with the soil using a mixer. In case of higher water contents than required, samples were left open at 4°C for a few hours to days until the desired water content was reached. After adjusting the water content, 75 g (dry weight) of soil was filled into plastic vessels, three replications each, and slightly compressed to a bulk density of 1.3 g cm⁻³ to create standardized conditions (Breulmann *et al.*, 2014). At least four blanks of empty plastic vessels were prepared for each incubation run. The samples were

then pre-incubated at 4°C for 120 h to level effects of mixing and water addition and to stabilize the respiration rate (Blagodatsky *et al.*, 2000).

Soil respiration measurements were conducted using an automated respirometer that allows incubating 95 samples in parallel (Respicond VIII, Nordgren Innovations AB, Sweden). The system provides a continuous measurement of CO_2 evolution by trapping CO_2 in potassium hydroxide (KOH) (Nordgren, 1988). The decrease in electrical conductivity in KOH solution caused by CO_2 entrapment was automatically measured every hour by platinum electrodes, and the changes in conductivity were automatically transformed to CO_2 evolution rates, based on equation 9.2 where A is a conductivity constant that depends on the molarity of the KOH solution, C_{t0} is the conductance of the fresh KOH measured at the beginning of the incubation and C_{t1} is the conductance at time t .

$$CO_2 = A * \frac{C_{t0} - C_{t1}}{C_{t0}} \quad (9.2)$$

Soil samples were sequentially set to 5°C, 10°C, 15°C, 20°C, and 25°C by heating the water bath of the Respicond system. Samples were kept at each temperature for 24 h (see also Gritsch *et al.*, 2015). The first 12 hours after each temperature rise were treated as equilibration time. This was necessary because soil microorganisms may need a couple of hours to adapt to the new temperature level. The proposed duration of this equilibration time differs among studies and mostly depends on the amount of soil sample used. Typically, they range from 2 h (Koch *et al.*, 2007) to 24 h (Vanhala *et al.*, 2008). Here, we decided for 12 h. The subsequent 12 hours were used for the calculation of soil respiration and were expressed as the average hourly CO_2 release. The average soil respiration rate from the three incubation vessels was used for further analyses. After completion of each temperature level, vessels were left open for about 30 minutes to equilibrate with ambient O_2 concentrations. The KOH solution was replaced subsequently. The short-term incubation approach was chosen to minimize effects of changing C pool sizes during the incubation. Longer incubation times can underestimate Q10 because SOC decreases with increasing incubation time (Hamdi *et al.*, 2013; Kirschbaum, 2006).

The ratio between soil respiration and SOC (SR_{25}/SOC ratio) was calculated as an indicator of SOC degradability (Craine *et al.*, 2010) from soil respiration at 45% WHC and 25°C (SR_{25}).

An exponential equation was used to calculate the relationship between temperature and soil respiration (Q10). The equation was fitted over the total temperature range of 5-25°C according to equation 9.3 where SR_T is soil respiration at a given temperature, a and b are fitted parameters, and T is temperature.

$$SR_T = a * \exp^{b*T} \quad (9.3)$$

The Q10 value was then calculated by inserting the parameter b into equation 9.4.

$$Q10 = \exp^{10*b} \quad (9.4)$$

9.2.5 Mid infrared spectroscopy (MIRS)

For MIRS analyses, sieved samples were dried at 40°C, milled, and about 20 mg was filled into microtiter plates and compacted with a plunger to create a plain and dense surface. Spectra were recorded using a Bruker Tensor 27 (Bruker HTS-XT), which records absorption from 7500 to 550 cm^{-1} wavenumber (corresponding to a range of 1333-18180 nm wavelength), with a resolution of 4 cm^{-1} . Spectra were automatically corrected for atmospheric water vapor and CO_2 . Each sample was replicated five times and represents an average of 120 scans. The average spectrum of the five replicates was used for further analyses and model construction.

9.2.6 Relation between MIRS spectra and soil respiration parameters

Pearson's product moment correlation coefficient between absorption at each wavenumber and soil respiration parameters (i.e., $SR_{25,45}$, Q10, SOC, and SOC-degradability) was calculated to get an impression, which individual spectral bands are indicative for the respective parameter. Correlation analyses were performed on baseline-corrected spectra in order to eliminate noise.

9.2.7 Partial least square regression (PLSR)

In a first step, we aimed at investigating the potential of MIRS to predict SR and the Q10 value under standardized conditions of temperature and soil moisture. Here, we chose SR at a temperature of 25°C and at a soil moisture level of 45% WHC ($SR_{25,45}$). Also for the determination of Q10 we used a WHC of 45%. A temperature of 25°C was chosen because it is often used as reference temperature in carbon models, e.g., in CLM, and a soil moisture level of 45% WHC was chosen because it represents an intermediate level. Further, we also tested the predictability of the SOC-normalized SR at 45% WHC and 25°C ($SR_{25,45}/SOC$), which is an indicator of SOC-degradability. This was done to exclude that the predictability of $SR_{25,45}$ only bases on a predictability of SOC contents.

The PLSR was performed using the OPUS QUANT software. The models were validated using leave-one-out cross validation, which is widely accepted especially for comparatively small sample sets (Bornemann *et al.*, 2008; Reeves *et al.*, 2001). Thus, $n - 1$ samples were used for model calibration and the performance of the model was validated by estimating the excluded sample. This procedure was repeated successively until each sample was excluded once. The number of ranks, the spectral preprocessing method, and the inclusion of spectral bands were selected in such a way that R^2 was maximized and the root mean square error of cross validation (RMSECV) was minimized. The selection of the optimal model was automatically performed by the optimization function of the OPUS QUANT software, which allows testing the performance of more than 1000 variants. PLSR for the prediction of Q10 and SR_{25} was performed on the complete data set and on subsets of individual land use types separately.

The prediction accuracy of MIRS-PLSR was evaluated by the coefficient of determination (R^2) and by residual predictive deviation (RPD), which is the ratio of the standard deviation of the reference data to the standard error of cross validation. The RPD is commonly classified as excellent for $RPD > 2.5$, good for RPD of 2.0-2.5, acceptable for RPD of 1.5-2.0, and poor for $RPD < 1.5$ (Cozzolino *et al.*, 2005; Viscarra Rossel *et al.*, 2007).

9.2.8 Random Forest modeling

For the simultaneous prediction of SR at any temperature and soil moisture level, a Random Forest model was used. Random Forest is a well-established and commonly used machine learning algorithm (Breiman, 2001a). It is known to be able to deal with nonlinearity, different types of predictor variables (e.g., nominal and metric data), complex interactions, as well as with collinearity of the predictors. Hence, Random Forest might be superior to PLSR analyses for more complex models. The algorithm bases on the concept of decision trees. As Random Forest is an ensemble method, several individual trees are built and the averaged predictions of the individual trees are taken as final estimate. For a detailed description of the algorithm see Breiman (2001a) and Kuhn and Johnson (2013a). Random Forest has also shown its applicability for MIRS based prediction models in Knox *et al.* (2015).

Random Forest modeling was performed in R using the implementation of Liaw and Wiener (2002) in conjunction with the wrapper package "caret" Kuhn (2017). As predictor variables, we included the MIRS spectra, the incubation temperature, soil moisture, and land use. The latter was necessary because the results from the MIRS-PLSR based predictions of Q10 values showed that different spectral bands are indicative for the Q10 value in each land use type. Random Forest was performed on baseline-corrected spectra and parts of the spectra were

removed before analysis (wavenumbers $>4000\text{ cm}^{-1}$) because they were assumed to be insensitive (see also Bornemann *et al.*, 2010). To investigate which variables were most important for the model accuracy, scaled variable importance was estimated according to Liaw and Wiener (2002).

The model performance and its ability to predict SR for unknown samples was estimated using a leave-one-sampling-point-out cross-validation. Each sampling point contains SR data for five temperatures and four soil moisture levels (i.e., 20 respiration measurements). Therefore, models were fitted by successively leaving the entire data of one sampling point out. The model was calibrated with the remaining respiration data (i.e., 2460 measurements), while the excluded data were used for model validation. This procedure was repeated successively until each sample was estimated with a model calibrated with all other samples. The leave-one-sampling-point-out approach was necessary as dependent data are prone to overfitting (Roberts *et al.*, 2017). This became obvious in a delusively high performance measure ($R^2 = 0.94$) when SR for a specific sample, temperature, and soil moisture level was estimated from a model, which included the same sample for calibration, but at other temperatures and soil moisture levels. Therefore, though Random Forest implements an internal performance indicator (out of bag error, Breiman, 2001a), it was necessary to exclude all data from the respective sample in view of a reliable validation.

9.3 Results and Discussions

9.3.1 MIRS-PLSR based prediction of soil respiration at standardized temperature and moisture

Soil respiration measured under standardized soil temperature (25°C) and moisture (45% of maximum water capacity), $SR_{25,45}$, was predictable with $R^2 = 0.83$ by MIRS-PLSR in a general model, which included the complete sample set (Figure 9.1a, Table 9.2). Splitting the data set into submodels of separate land use types revealed no considerable improvement of RPD or R^2 (9.1b-d, Table 9.2).

It is obvious that the predictability of $SR_{25,45}$, did not originate from a direct correlation between the MIRS spectra and soil respiration. Actually, no respiration takes place in the dried samples during the MIRS measurement. Soil respiration was linked to the MIRS spectra through surrogate correlations, i.e., through correlation of $SR_{25,45}$ to SOC content, SOM quality or other soil parameters that are reflected in the infrared spectra and which control SR. The driving surrogate parameter was likely the SOC content, first, because it directly affected soil respiration (here: $R^2 = 0.59$, $p < 0.01$ for the relation between SOC and SR),

and second, because it is known to be accurately predictable by MIRS (here: $R^2 = 0.95$, $RPD = 4.36$, see also Bornemann *et al.*, 2010; Reeves, 2010; Zimmermann *et al.*, 2007). Indeed, $SR_{25,45}$ was correlated to similar wavenumbers of the MIRS spectra as the SOC content (Fig. 9.2). Also Chang *et al.* (2001) reported for a near infrared range that spectral predictors of SR were similar to those of SOC.

At a first glance, such a surrogate correlation questions the need for MIRS-PLSR to predict spatial patterns of SR because SOC contents can be measured with similar or even less efforts of time and costs by elemental analysis. However, linear regressions between SOC contents and $SR_{25,45}$ explained only 59% of data variability, while MIRS-PLSR based predictions of $SR_{25,45}$ were more accurate ($R^2 = 0.83$). Consequently, other surrogate parameters than SOC content contributed to the MIRS-based prediction of SR. Indeed, when normalizing SR to SOC contents ($SR_{25,45}/SOC$), the cross validation revealed that MIRS captured at least parts of the SOC degradability, especially in grasslands and forests (Table 9.2). This suggests that MIRS did not only account for quantitative but also for qualitative properties of SOC, as already discussed by Bornemann *et al.* (2010) and Ludwig *et al.* (2008). Indeed, soils with high $SR_{25,45}$ revealed high absorptions in spectral regions typically assigned to aliphatic C-H stretches (Fig. 9.2, Fig. 9.3, 2925 cm^{-1} and 2850 cm^{-1} , Rumpel *et al.*, 2001), which were probably associated with fresh and labile C (Margenot *et al.*, 2015). Positive correlations between absorption and $SR_{25,45}$ were further found between 1100 cm^{-1} and 1200 cm^{-1} , associated with the C-O stretch of polysaccharides (Fig. 9.2, Fig. 9.3, Rumpel *et al.*, 2001), at 1660 cm^{-1} (carbonyl-C; Rumpel *et al.*, 2001), as well as at 1722 cm^{-1} (C=O stretching of COOH; Rumpel *et al.*, 2001). Not all components in organic matter promote respiration. Indeed, negative correlations to SOC and $SR_{25,45}$ appeared, for instance, in the region between 1250 cm^{-1} and 1615 cm^{-1} , which Rumpel *et al.* (2001) assigned to aromatic species. The latter are usually less prone to decomposition Zech *et al.* (1992). Hence, it is likely that both the quantity of SOC and the composition of soil organic matter affected the MIRS spectra and the prediction of $SR_{25,45}$.

9.3.2 MIRS-PLSR based prediction of Q10 values

The MIRS-PLSR based prediction of Q10 values was not successful for the entire data set ($R^2 = 0.43$; Fig. 9.4a, Table 9.2). Several studies reported that the predictive power can be enhanced by splitting the data set into subsets to reduce the heterogeneity within the data set (Bornemann *et al.*, 2010; Linker *et al.*, 2006). Thus, we divided the data set according to land use types and calibrated separate submodels. Within the cropland and the grassland submodel, Q10 values were predictable with $R^2 = 0.72$ and $R^2 = 0.75$, respectively (Fig. 9.4, Table, Table 9.2). The Q10 value of forest soils was not predictable by MIRS-

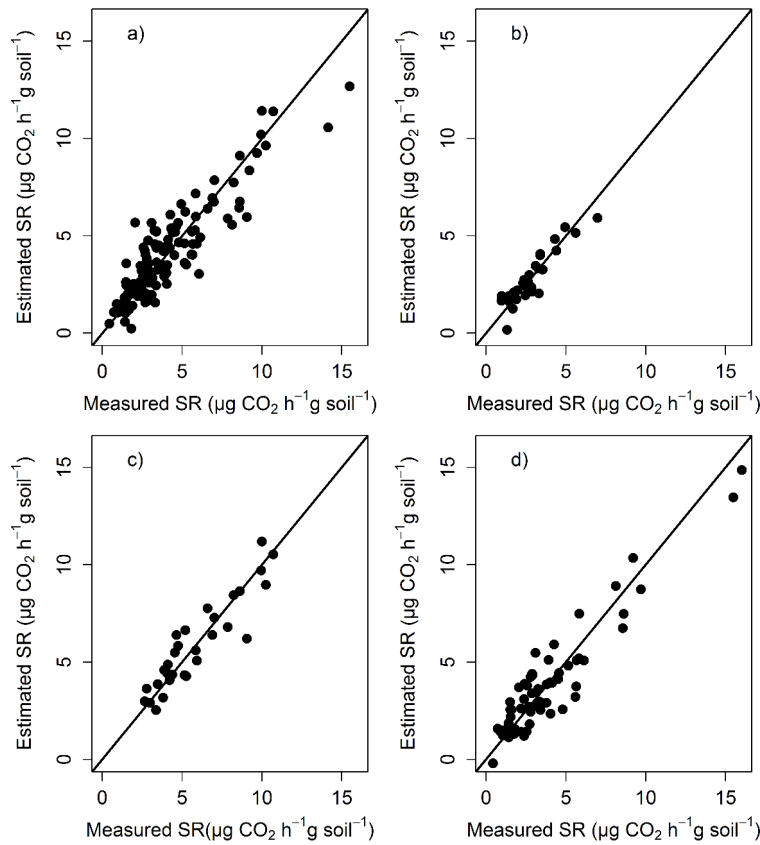


Figure 9.1: Comparison between measured and predicted soil respiration (SR) rates at 25°C and 45% of water holding capacity based on leave-one-out cross validation. a) general model, b) cropland submodel, c) grassland submodel, d) forest submodel.

PLSR ($R^2 = 0.03$, Fig. 9.4d, Table 9.2).

The root mean square error of cross validation (RMSECV) of the cropland and grassland submodels was always lower than the RMSECV derived from the general model for the respective cropland and grassland samples (supplementary information: Table S1). Hence, the comparatively poor accuracy of the general model did not solely result from the poor predictability of forest soils. Even the cropland and grassland soils, which were successfully predicted by the respective submodel, were less accurately predicted by the general model (Fig. 9.4a). In line with Meyer *et al.* (under review) and Zheng *et al.* (2009), regulating factors of Q10 values varied among land use types. Hence, the MIRS-based prediction of Q10 values might rely on different surrogate correlations within each land use type. This might complicate the accurate prediction of Q10 values in a general model. Thus, we recommend using land use specific submodels for MIRS-PLSR based predictions of Q10 values.

Q10 was correlated to the same spectral regions that were indicative for SOC degradability (i.e., the SOC-normalized SR, $SR_{25,45}/\text{SOC}$), but inversely. Spec-

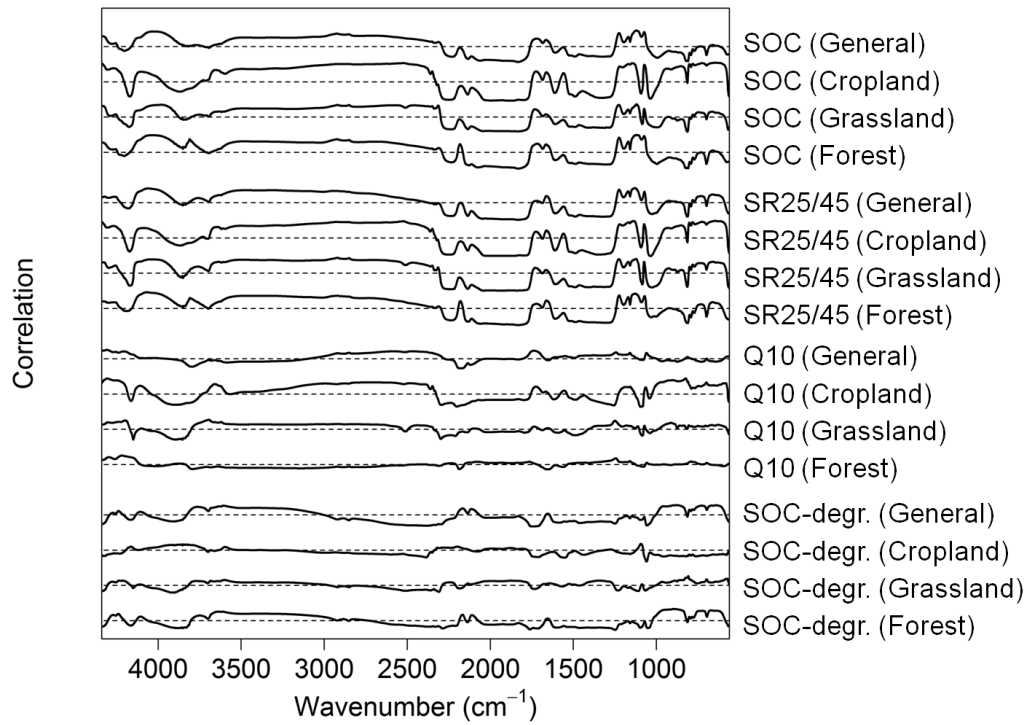


Figure 9.2: Correlation between absorption at each wavenumber and SOC, SR_{25,45}, Q10, and SOC-degradability for the general dataset and the individual subsets. The correlation between absorption and soil parameter is zero when the solid line matches the dashed line. Correlations are positive for values above the dashed line and negative below the line. SOC = soil organic carbon, SR_{25,45} = soil respiration measured at 25°C and 45% of water holding capacity, Q10 is the temperature sensitivity, i.e., the factor for the increase in soil respiration by a 10°C rise of temperature, SOC-degr = SOC degradability as derived from the SOC normalized soil respiration rate, i.e., SR_{25,45}/SOC.

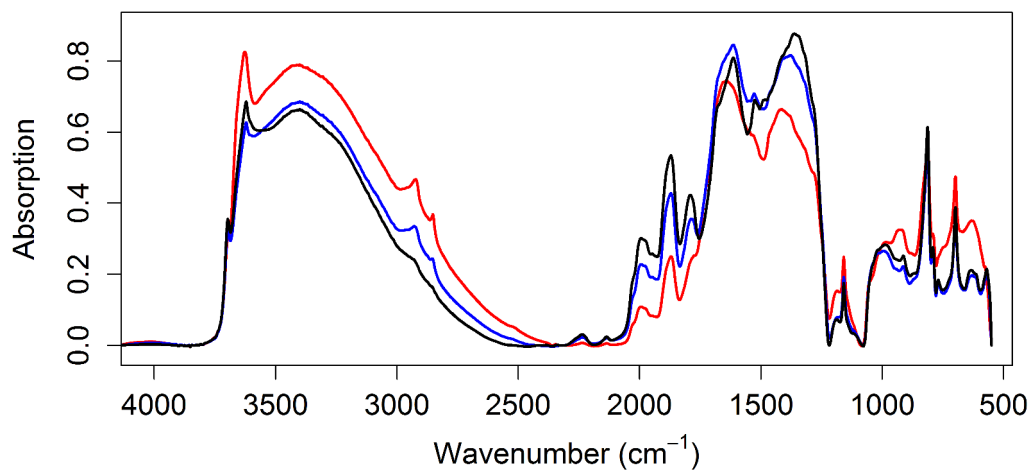


Figure 9.3: Baseline-corrected MIRS spectra of three grassland soils with different SR_{25,45} rates of 0.97 μg CO₂ h⁻¹ g soil⁻¹ (black line), 5.86 μg CO₂ h⁻¹ g soil⁻¹ (blue line), and 10.25 μg CO₂ h⁻¹ g soil⁻¹ (red line).

Table 9.2: Prediction accuracy for $SR_{25,45}$, Q10 values, and C-degradability (i.e., $SR_{25,45}/SOC$, SOC normalized soil respiration) based on the different sub-models. ^anumber of samples, ^b R^2 = coefficient of determination, ^cRMSECV = Root Mean Squared Error of cross-validation, ^dRPD = residual predictive deviation

Parameter	Model	na	Spectral preprocessing	R^2 ^b	RMSECV ^c	RPD ^d	Rank
$SR_{25,45}$	General	124	First derivative	0.83	1.2	2.45	6
	Crop	29	No preprocessing	0.83	0.565	2.45	9
	Grass	30	Multiplicative scattering correction	0.84	0.951	2.52	5
	Forest	63	Second derivative	0.87	1.08	2.8	7
Q10	General	124	No preprocessing	0.43	0.234	1.32	10
	Crop	29	Second derivative	0.72	0.135	1.91	10
	Grass	30	Straight line subtraction	0.75	0.111	2.0	10
	Forest	63	Straight line subtraction	0.03	0.244	1	1
C-degradability ($SR_{25,45}/SOC$)	General	123	Second derivative	0.48	0.529	1.38	8
	Crop	30	Second derivative	0.07	0.394	1.04	9
	Grass	30	Vector normalization (SNV)	0.74	0.24	1.96	4
	Forest	62	Straight line subtraction	0.68	0.242	1.77	8

tral regions that correlated positively with SOC degradability correlated negatively with the Q10 value and vice versa. This observation is in line with the C-quality-temperature hypothesis (Bosatta and Agren, 1999; Conant *et al.*, 2008; Lefèvre *et al.*, 2014). According to this hypothesis, SOC that is easily degradable is less sensitive to temperature changes than recalcitrant SOC. Only in forest soils, no spectral region was correlated with the Q10 value, thus supporting previous indications that Q10 variability in forest soils could not be assigned to any measured physicochemical soil properties or to C-degradability (Meyer *et al.*, under review).

Altogether, our results showed that MIRS-PLSR is potentially suitable to predict Q10 values but that its applicability is limited to cropland soils and grassland soils. The models could be used to derive an approximation of Q10 values but should not be used to replace conventional respiration measurements when highly resolved data are required.

9.3.3 Simultaneous prediction of soil respiration across various levels of soil moisture and temperature by Random Forest modeling

Although the predictability of SR at a reference temperature (e.g., 25°C) and of Q10 values allows calculating SR at any other temperature based on equation 9.1, the above presented results were only valid for a single soil moisture level (i.e., here 45% of WHC). Although the prediction of SR_{25} was also possible for other soil moisture levels (supplementary information: Table S2), quite a lot of separate MIRS-PLSR models would be required. Further, also the Q10 value is

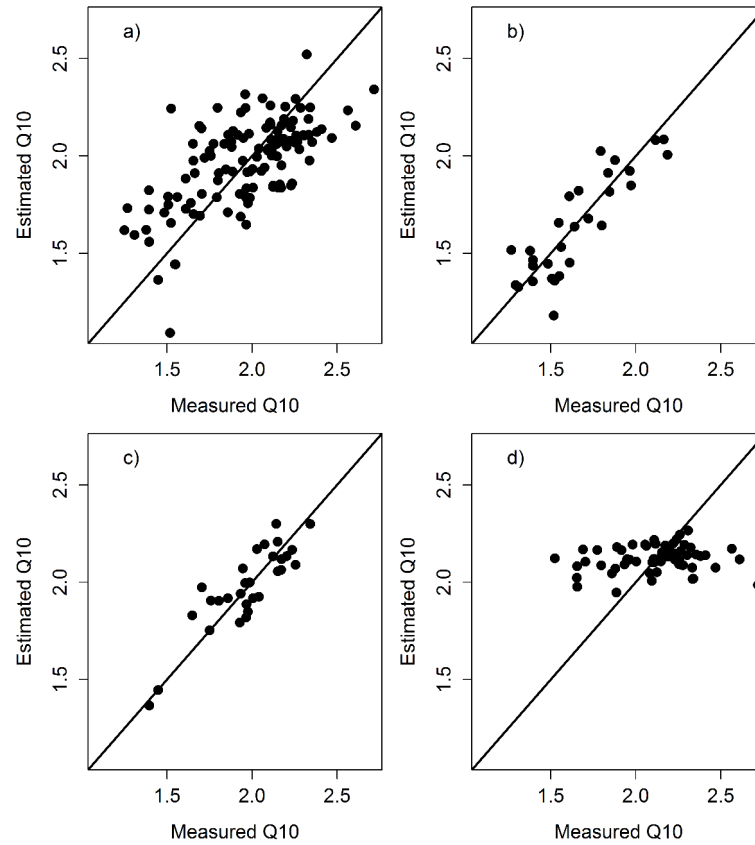


Figure 9.4: Comparison between measured and predicted Q_{10} values based on leave-one-out cross validation. a) general model, b) cropland submodel, c) grassland submodel, d) forest submodel.

generally assumed to vary across soil moisture levels (e.g., Craine and Gelderman, 2011; Luan *et al.*, 2013; Meyer *et al.*, under review). While the above presented models might be adequate for certain purposes, i.e., to study spatial patterns of SR or Q_{10} at a single soil moisture level, a model that allows for the simultaneous prediction of SR at any temperature and soil moisture level would be of great help to estimate SR under various weather conditions.

Thus, we tested if SR at any given temperature and soil moisture level can be simultaneously predicted by a Random Forest model. Here, besides the MIRS spectra, also temperature, soil moisture, and land use were used as predictor variables of SR. By introducing land use as predictor variable, the Random Forest model might consider that different surrogate correlations, i.e., specific spectral regions, are required for each land use type. This was needed to account for the temperature sensitivity of SR (section 9.3.2). Due to the addition of land use to the predictor variables, no land use specific submodels were required.

The Random Forest model allowed to predict SR at any given soil moisture level and temperature with an R^2 of 0.73 (Fig. 9.5). The SR of croplands, grasslands, and forests was well represented by the general model (Fig 9.6). The

most important variable for the model performance was temperature, followed by moisture, and land use (Fig 9.7). Further, five different wavenumbers turned out to be important for model calibration: 1672 cm^{-1} , 586 cm^{-1} , 3998 cm^{-1} , and 1840 cm^{-1} , and 1838 cm^{-1} . These importance rankings were supported by simple correlation analyses where the concerning wavenumbers showed high positive or negative correlations with SR (Fig 9.2).

In summary, our results revealed that the inclusion of MIRS into SR estimates indirectly adds information on SOC content, SOC-degradability, and temperature sensitivity to the predictor variables. Hence, we assume that the MIRS based prediction has a large advantage over approaches that predict SR solely on the basis of temperature and soil moisture (e.g., Bowden *et al.*, 1998; Keith *et al.*, 1997). With respect to the consideration of several surrogate parameters, the large range of considered soils, the fast and inexpensive measurement, and the strict validation method (see section 9.2.8), we are convinced that the MIRS-based Random Forest model has a large potential regarding the range of applicability and the accuracy to predict SR for unknown samples.

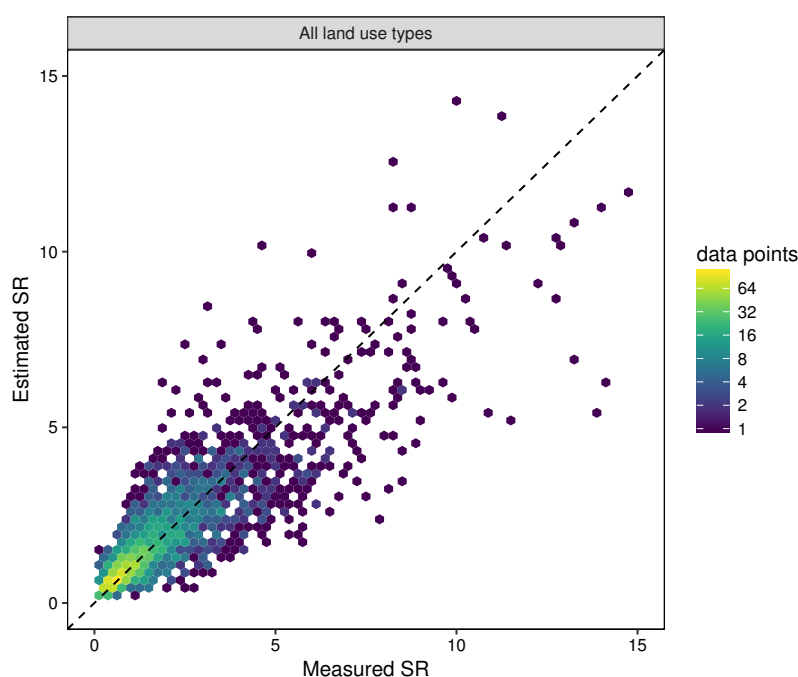


Figure 9.5: Comparison between measured and predicted soil respiration rates based on leave-one-sampling-point-out cross-validation. Measured respiration rates from five temperatures and four soil moisture levels per sample were included. For an easy visual interpretation, the data are presented via hexagon binning where the numbers of data points, which are located in each hexagon, are depicted by color.

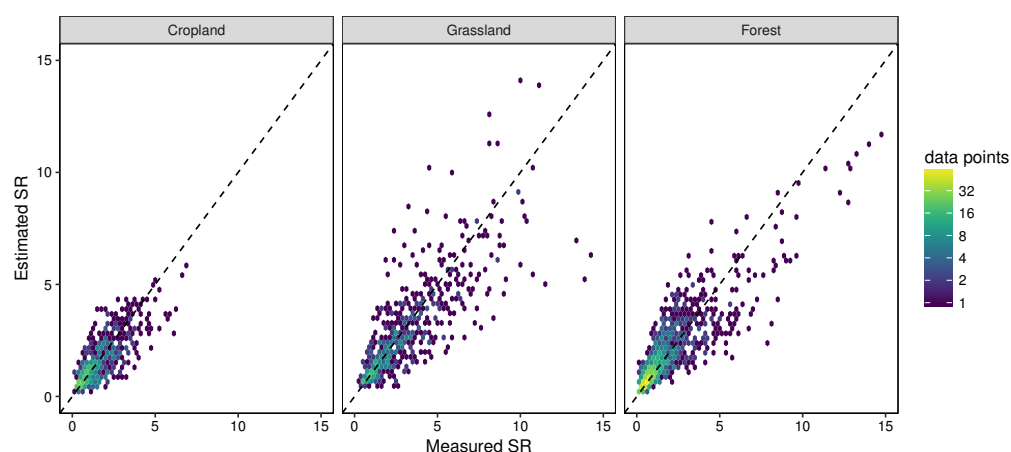


Figure 9.6: Comparison between measured and predicted soil respiration rates based on leave-one-sampling-point-out cross-validation for each land use type separately. a) cropland soils within the general model, b) results for grassland soils within the general model, c) results for forest soils within the general model. Note that results were derived from the general model to show the differences in model performance for each separate land use type. No submodels were calculated. For an easy visual interpretation, the data are presented via hexagon binning where the numbers of data points, which are located in each hexagon, are depicted by color.

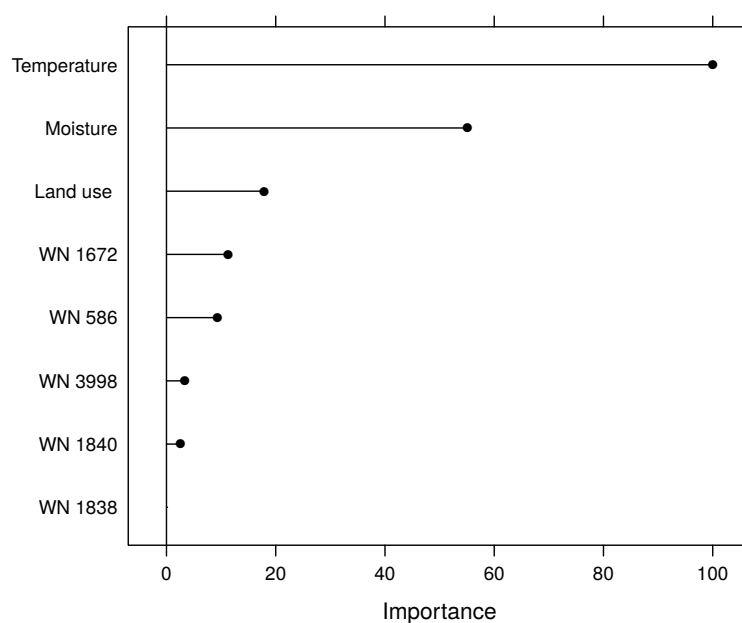


Figure 9.7: Scaled variable importance as revealed by the Random Forest algorithm. The higher the importance score, the more important is the variable for the soil respiration estimate. WN = wavenumber.

9.4 Conclusion

In this study, we showed that MIRS combined with sophisticated statistical methods represents a promising tool for the spatiotemporal prediction of soil respiration. The MIRS models should not replace conventional lab analyses when high accuracy data are required, however, they could be used in studies where lab analyses are not feasible (e.g., for large data sets), and they appear useful in replacing simple pedotransfer functions that rely solely on, e.g., SOC contents. The developed Random Forest model offers a new possibility for a high throughput estimate of SR patterns across the landscape and for various temperatures and soil moisture levels at least within the Rur catchment. An extension of the database to soils from other regions is now required.

Acknowledgments This work was supported by the collaborative research centre SFB/TR32 "Patterns in Soil-Vegetation-Atmosphere Systems: Monitoring, Modeling, and Data Assimilation" funded by the Deutsche Forschungsgemeinschaft (DFG). We thank Peter Grässler, Jana Bolduan, and Carmen Meier for their support in laboratory work.

Chapter 10

Regional-scale controls on the spatial activity of rockfalls (Turtmann valley, Swiss Alps) – A multivariate modelling approach

Karoline Messenzehl (1), Hanna Meyer (2), Jan-Christoph Otto (3),
Thomas O. Hoffmann (1,4), Richard Dikau (1)

(1) Department of Geography, University of Bonn, Bonn, Germany

(2) Faculty of Geography, Philipps-University Marburg, Marburg, Germany

(3) Department of Geography and Geology, University of Salzburg, Salzburg, Austria

(4) Federal Institute of Hydrology, BfG, Koblenz, Germany

Published in

Geomorphology 2017, 287, 29–45

Received 1 July 2015

Revised 15 January 2016

Accepted 19 January 2016

Available online 22 January 2016

Published 15 June 2017

10 Regional-scale controls on the spatial activity of rockfalls (Turtmann valley, Swiss Alps) – A multivariate modelling approach

Karoline Messenzehl, Hanna Meyer, Jan-Christoph Otto, Thomas Hoffmann, Richard Dikau

Abstract

In mountain geosystems, rockfalls are among the most effective sediment transfer processes, reflected in the regional-scale distribution of talus slopes. However, the understanding of the key controlling factors seems to decrease with increasing spatial scale, due to emergent and complex system behaviour and not least to recent methodological shortcomings in rockfall modelling research. In this study, we aim (i) to develop a new approach to identify major regional-scale rockfall controls and (ii) to quantify the relative importance of these controls. Using a talus slope inventory in the Turtmann Valley (Swiss Alps), we applied for the first time the decision-tree based random forest algorithm (RF) in combination with a principal component logistic regression (PCLR) to evaluate the spatial distribution of rockfall activity. This study presents new insights into the discussion on whether periglacial rockfall events are controlled more by topo-climatic, cryospheric, paraglacial or/and rock mechanical properties.

- (i) Both models explain the spatial rockfall pattern very well, given the high areas under the Receiver Operating Characteristic (ROC) curves of > 0.83 . Highest accuracy was obtained by the RF, correctly predicting 88% of the rockfall source areas. The RF appears to have a great potential in geomorphic research involving multicollinear data.
- (ii) The regional permafrost distribution, coupled to the bedrock curvature and valley topography, was detected to be the primary rockfall control. Rockfall source areas cluster within a low-radiation elevation belt (2900-3300 m a.s.l.) consistent with a permafrost probability of $> 90\%$. The second most important factor is the time since deglaciation, reflected by the high abundance of rockfalls along recently deglaciated (< 100 years), north-facing slopes. However, our findings also indicate a strong rock mechanical control on the paraglacial rockfall activity, declining either exponentially or linearly since deglaciation.

The study demonstrates the benefit of combined statistical approaches for predicting rockfall activity in deglaciated, permafrost-affected mountain valleys and highlights the complex interplay between rock mechanical, paraglacial and

topo-climatic controls at the regional scale.

Keywords Rockfall; Logistic regression model; Random Forests algorithm; Rock mechanical properties; Paraglacial; Permafrost

10.1 Introduction

Rockfalls from steep rock slopes are frequent in cold mountain environments (Rapp, 1960; Whalley, 1984), where they represent a considerable natural hazard (Hungre *et al.*, 1999; Ravanel and Deline, 2015). From a geomorphological perspective, detailed knowledge of the key factors controlling rockfalls is important as rockfalls efficiently contribute to bedrock denudation (Heim, 1932; Selby, 1982a; Krautblatter *et al.*, 2012), accumulate massive talus deposits at the footslopes (Caine, 1974; Messenzehl *et al.*, 2014) and play a major role in the sediment flux in mountain geosystems (Jäckli, 1957; Hoffmann *et al.*, 2013; Heckmann *et al.*, 2016). The spatial occurrence of rockfalls, i.e. covering debris falls ($<10 \text{ m}^3$), boulder falls ($10\text{--}102 \text{ m}^3$) and block falls ($102\text{--}104 \text{ m}^3$) (nomenclature after Krautblatter *et al.* (2012), is best indicated by the presence of talus slopes (Hales and Roering, 2005; Sass, 2005c; Moore *et al.*, 2009) integrating a long-term average of rockfall history over longer time scales such as the Holocene (Krautblatter and Dikau, 2007). However, the heterogeneous spatial distribution of talus landforms within mountain landscapes as well as their varying volumes and material properties (Gerber, 1974; Schrott and Hoffmann, 2003; Sass, 2010) reflect a complex interplay of numerous causative factors, collectively defining the sensitivity of rockwalls to fail at different spatial and temporal process scales (Fig. 10.1). At small scales, the balance between shear stresses and shearing resistance of the rock mass (Terzaghi, 1962) is determined by the highly spatio-temporal variable interplay between mechanical, thermal and hydrological bedrock characteristics (Hoek and Brown, 1997; Wyllie and Mah, 2004) as well as rock breakdown processes (e.g. Dixon and Thorn (2005); Gunzburger *et al.* (2005); Matsuoka and Murton (2008); Hall and Thorn (2014)). Rockfalls are particularly found to correlate directly with a low intact rock strength (Selby, 1980; Vehling *et al.*, 2016), an unfavourable joint orientation in relation to slope surface (Cruden and Hu, 1994; Moore *et al.*, 2009) and a high joint density (Sass, 2005b; Loye *et al.*, 2012). Sheeting joints parallel to the slope surface are often associated to stress relaxation after glacial unloading of rock slopes during paraglacial conditions (Brunner and Scheidegger, 1973; Augustinus, 1992). Although the synergy of multiple weathering processes is increasingly acknowledged (Hall, 2006; Hall *et al.*, 2012; Viles, 2013), daily freeze-thaw cycles (Matsuoka, 1994; Sass, 1998) and seasonal or multi-annual segregation ice growth are supposed to be the prime destabilising agents in cold environments (Hallet *et al.*, 1991; Matsuoka and Sakai, 1999; Murton *et al.*, 2006). The efficiency of freeze-thaw action significantly depends on bedrock moisture (Coutard and Francou, 1989; Prick, 1997) and the thermal behaviour of permafrost (Allen *et al.*, 2009; Krautblatter *et al.*, 2013). Likewise, the specific bedrock roughness and morphometry, caused e.g. by convex overhangs, can lead to spatially variable, but persistent stress fields (Gerber and

Scheidegger, 1969).

With increasing scale, the spatio-temporal variability of most rockfall controls generally decreases pre-disposing the rock slope to fail (Crozier, 1989). Key factors seem to be primary geology and topo-climate (Fig. 10.1) considering that instabilities predominantly occur on cold, shaded slopes, steeper than 40° (Dorren and Seijmonsbergen, 2003; Wichmann *et al.*, 2009). These sites are often congruent with an altitudinal belt of low solar radiation (Noetzli *et al.*, 2003; Fischer *et al.*, 2012), where frost cracking over the active permafrost layer might be highly effective (Gruber, 2005; Hales and Roering, 2009). Over geological time scales, the lithological and tectonic settings cause an inherent preconditioning of rockwalls for failures (Cruden and Hu, 1998; Coe and Harp, 2007). A particular bedrock pre-disposition is linked to the impact of glaciation during LGM due to slope oversteepening as well as subsequent paraglacial adjustment processes including slope debuttressing, stress release, permafrost degradation and isostatic rebound (Ballantyne, 2002; McColl, 2012; Leith *et al.*, 2014).

Despite our knowledge on the process scale of rockfall causative factors (Fig. 10.1), their relative importance within their complex interplay is insufficiently understood. A major challenge is that the relative roles of rockfall controls probably changes depending on scale. According to Harrison (2001) and Phillips (1988, 2003) it must be supposed that the local-scale causes for failure of individual rock slopes likely contrast with those being causative at the regional scale of mountain catchments due to emergent system behaviour and increasing system complexity (and often non-linearity) with increasing spatial scale (Fig. 10.2B, C-D). Harrison (2001) even assumes that at each scale, new, often unknown system properties and causalities emerge (“?” in Fig. 10.2D) being insensitive to changes at lower levels of this hierarchical structure (see also de Boer, 1992; Church, 1996). However, compared to the slope scale, the understanding of bedrock destabilising factors at regional scales is still very limited as shown by the recent debate on whether rockfall activity is dominated either by topo-climatic forcing (Tricart and Cailleux, 1972; Büdel, 1977; Hales and Roering, 2005), paraglacial adjustment (Cossart *et al.*, 2014; Feuillet *et al.*, 2014) or rock mechanical properties (Duarte and Marquinez, 2002; Fischer *et al.*, 2006).

To some extent, this limited systemic knowledge may be due to methodological shortcomings. While most work focuses on local rock instabilities, few studies have examined regional-scale rockfall controls so far (e.g. Duarte and Marquinez, 2002; Hales and Roering, 2005; Frattini *et al.*, 2008; Michoud *et al.*, 2012). We argue that the appropriate research design to identify key rockfall controls is dictated by the specific scale of interest. While reductionist, process-based field surveys are appropriate to study local parameters (Krautblatter and Moore, 2014), the complex and emergent behaviour of destabilising factors at a larger scale

might be best understood by abductive approaches (Peirce, 1902) using e.g. inventories of talus slopes (e.g. Moore *et al.*, 2009; Tanarro and Muñoz, 2012). Here, GIS-based multiple logistic regressions provide a powerful statistical method that has been successfully applied in the mapping of permafrost (Janke, 2005) and patterned grounds (Miska and Jan, 2005), soil erosion (Vanwallegghem *et al.*, 2008) and extensively of landslide susceptibility (e.g. Guzzetti *et al.*, 1999; Ohlmacher and Davis, 2003; Vanacker *et al.*, 2003; Bai *et al.*, 2010; Borgomeo *et al.*, 2014). More recently, machine learning algorithms such as random forests (Breiman, 2001a), mainly known from ecological studies (Cutler *et al.*, 2007) and climate modelling (Meyer *et al.*, 2016b; Kühnlein *et al.*, 2014b) are receiving increased attention in landslide studies (Brenning, 2005; Stumpf and Kerle, 2011; Vorpahl *et al.*, 2012; Catani *et al.*, 2013). However, since knowledge of deep-seated gravitational mass movements cannot be automatically transferred to rockfall processes, adequate approaches are needed specifically for rockfall research.

To address the recent systemic and methodical shortcomings, we investigate the spatial rockfall activity in the Turtmann Valley (Swiss Alps). Our objectives are (i) to develop an appropriate approach to identify major factors controlling the regional-scale occurrence of rockfall processes and (ii) to evaluate the relative importance of rock mechanical, paraglacial and topo-climatic influences within this complex interplay. Using an inventory of 220 talus slopes, we combine a classical logistic regression model based on principal components and a novel random forests classification to examine the relative causality between the slopes' rockfall susceptibility and ten potential destabilising parameters.

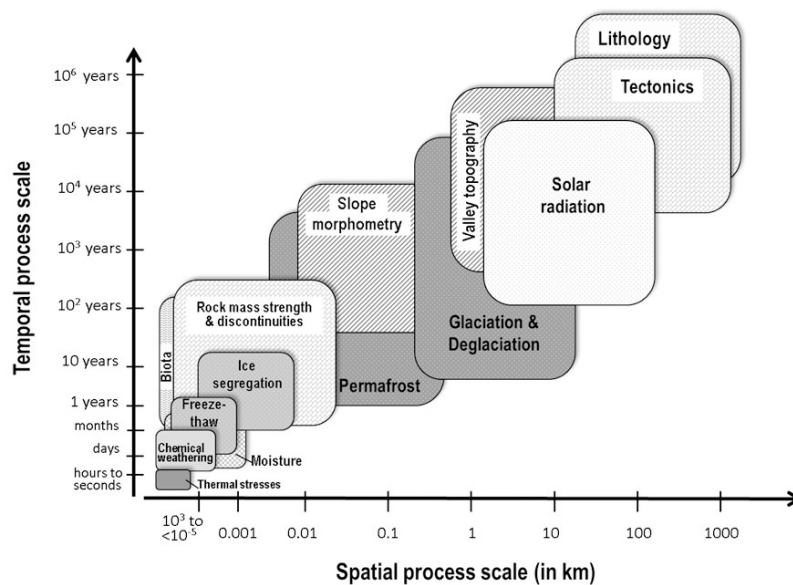


Figure 10.1: Process-scale of potential rockfall controls with respect their temporal and spatial variability. So far, limited knowledge exists on the relative importance of topo-climatic, morphometric, paraglacial and rock mechanical factors for the regional-scale rockfall activity.

10.2 Characteristics of the study area

The Turtmann Valley is located in the Valais Alps (Switzerland) between the Matter Valley and the Anniviers Valley (Fig. 10.2), covering a catchment area of 110 km² at altitudes ranging from 620 m to 4200 m above sea level (a.s.l.) (Otto *et al.*, 2009). The 15 km long valley is a typical Pleistocene glacial trough with the Turtmann Glacier at the valley head. While the rock slopes of the trunk valley are mainly affected by deep-seated gravitation processes, rockfalls predominate in the 14 strongly W-E-oriented hanging valleys, lying on either side of the trough shoulders at 2300-2600 m a.s.l.

The lithology of the hanging valleys is dominated by metamorphic rocks (penninic Siviez-Mischabel nappe) consisting of Palaeozoic micashists and paragneisses (ca. 72%), which form most of the bedrock in the northern and eastern hanging valleys located in the NE (Labhart, 2009). Thin layers of amphibolite (0.20%), quartzites (0.53%) and apatite (6.23%) are incorporated. As result of the tectonic folding, the metamorphic rocks generally strike in a south-west direction with a dip of 20-30° (Bearth, 1980). Overlying the crystalline rocks, mesozoic dolomites, limestones and marbles with clay layers Barrhorn series and Frilhorn series) occur in the western (Meidtälli) and south-eastern hanging valleys (Pipjitälli) (Labhart, 2009). Due to the inner alpine location, the climatic situation of the study area is characterised by dry continental conditions with mean annual precipitation of ca. 600-900 mm at 2000 m a.s.l. (Gärtner-Roer *et al.*, 2013) and a snow line at ca. 3450 m (Escher, 1970). The 0°C isotherm of the mean annual air temperature might be at ca. 2550 m a.s.l. (van Tatenhove and Dikau, 1990). According to a local permafrost model (Nyenhuys *et al.*, 2005), 37 km² or 33% of the catchment area is very likely affected by permafrost (> 60% probability), with a lower limit ranging from 2500 m a.s.l. (N-orientations) to 3000 m a.s.l. (S-orientations).

The ice surface of the Late-Glacial maximum (LGM) is supposed to have reached up to 2600 m a.s.l. in the main valley, rising towards the cirques of the hanging valleys and the Bishorn peak (4058 m) (Kelly *et al.*, 2004). At ca. 24-21 kyrs BP, rock slopes were likely ice covered by local hanging glaciers and only the peaks were ice-free nunataks. After ca. 18 kyrs BP, deglaciation started in the main trough and subsequently in the hanging valleys (Kelly *et al.*, 2004; Schlüchter, 2004), successively exposing the slopes above the trough shoulders. However, no data on the timing of glacier recession in the study area exist. With respect to the Late Glacial advance, the Younger Dryas glacier (ca. 12-8 kyrs) probably did not affect the rock faces of the hanging valleys (Ivy-Ochs *et al.*, 2009) indicated by the well-preserved Egesen (Younger Dryas) moraines within the hanging valleys at ca. 2600 m (Otto and Dikau, 2004). Today, the few glaciers

still existing in some of the hanging valleys are affected by a significant recession trend due to atmospheric warming in the last decades.

Rockfall events of small ($< 10 \text{ m}^3$) and medium magnitude ($10\text{--}104 \text{ m}^3$) (Messenzehl and Draebing, 2015) are some of the most active processes in the hanging valleys leading to a postglacial denudation rate of around $0.7\text{--}2.2 \text{ mm/a}$ (Otto *et al.*, 2009). A total number of 220 active talus slopes (including sheets and cones) accounts for ca. $8.7\text{--}12.3\%$ of the tributaries' total sediment volume (Otto *et al.*, 2009), representing major sources for rock glaciers and debris flows in the sediment cascade. Sediment output from the tributaries to the main drainage system is largely disconnected due to the bedrock trough shoulders and geomorphic buffers such as moraine landforms (compare with Messenzehl *et al.*, 2014).

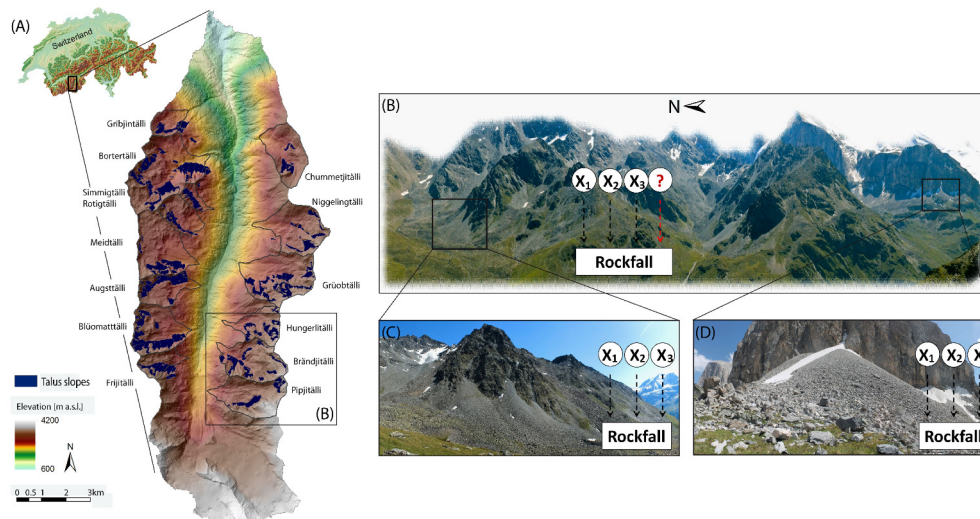


Figure 10.2: Study area. A total of 220 Talus slopes (mapped in blue) are deposited in the 14 hanging valleys of the Turtmann Valley, southern Swiss Alps (A). It is supposed that the relative role of regional-scale rockfall controls (X_1 , X_2 , X_3 , ?) (B) contrasts to destabilising factors (X_1 , X_2 , X_3) at the local scale (C, D) due to emergent and complex behaviour of geosystems with increasing spatial scale.

10.3 Modelling approach

10.3.1 Data selection and pre-processing

10.3.1.1 Response variable

An inventory of 220 talus slopes (Otto *et al.*, 2009) was used to deduce abductively the rockfall activity of the associated rock slope. Since rockfalls predominantly occur in the hanging valleys, the trunk valley was excluded from

further investigations. The presently absent or negligible vegetation cover on talus deposits indicates active rockfalls. Rockfall source areas in bedrock were automatically extracted from a 1 m HRSC-Digital Elevation Model (DEM, Otto *et al.*, 2007) using the hydrological algorithm of the D8 flow routing in SAGA GIS (Martz and Garbrecht, 1992). The talus slopes served as sinks to determine their contributing areas. Our approach is in accordance with Frattini *et al.* (2008) and Marquínez *et al.* (2003) similarly using inventories of talus slopes to deduce associated grid cells of rockfall source area polygons. To derive rockfall initiation zones exclusively on bedrock, we furthermore intersected the contributing areas with the bedrock outcrop derived from a digital geological map 1:25000 (© Swiss Topo, based on Bearth, 1980). To correct the planimetric area of the rockfall source areas for the actual surface area, each raster cell was multiplied by the cosine of the slope gradient. For subsequent modelling, source areas were transformed into a binary raster grid (1 = presence of rockfall source area, 0 = no rockfall source area) and used as dependent response variable. A ground truthing based on field observations (e.g. fresh rockfall detachment zones, or freshly weathered bedrock) and interpretation of HRSC-aerial images (ground resolution 50cm) suggests reasonable results of the automatic mapping.

10.3.1.2 Predictor variables

Based on systemic process understanding and data availability, ten variables were considered to be potentially causative for the spatial rockfall distribution (Table 10.1). Topographic and morphometric parameters including elevation, slope gradient (Fig. 10.3A), aspect, profile curvature and surface roughness (Fig. 10.3B) were derived from the 1 m DTM using ArcGIS (ESRI, 2006). Considering climatic and cryospheric variables (Fig. 10.3C,D), the potential annual sum of solar radiation in watt hours per square meter (WH/m²) was modelled for every pixel using a hemispherical viewshed algorithm provided in ArcGIS (Fu and Rich, 2002). A layer of permafrost probability was derived from the empirical-statistical model PSIM (Permafrost Simulation Indication Model) by Nyenhuis (2006). Here, a local inventory of active and relict rock glaciers is related to potential direct solar radiation and mean annual air temperature (represented by altitude) (Nyenhuys *et al.*, 2005). In contrast with other permafrost models, shadowing effects due to relief and solar variations are considered.

Rock mechanical characteristics including lithology and overall joint orientation were extracted from the digital geological vector map at 1:25000 scale (© Swiss Topo, based on Bearth, 1980). Six lithological classes (Fig. 10.3E) were identified: 1) marble and limestone, 2) paragneiss and micaschist, 3) amphibolite, 3) apatite, 4) quartzite and 5) basalt. Quaternary deposits are not included. Additionally, joint orientation (dip and dip direction) was extracted for those rock

Table 10.1: *Overview on response and predictor variables with respect to geomorphic type, data source, variable scale, decoding scheme (classes with rockfall density RD) and their implication for slope (in)stability based on a selection of related rockfall and rockwall instability studies.*

Type and name of variable	Data Source	Variable scale	Implication for slope stability (referring to a selection of local- and regional-scale rockfall and rockwall instability studies)	
			Response variable (dependent variable)	Decoding
Rockfall source areas	Talus slope inventory (Otto <i>et al.</i> , 2009)	binary	-	-
Predictor variable (independent variables)				
Topography	Elevation Slope Aspect	DEM (1 m) DEM (1 m) DEM (1 m)	metric ordinal ordinal	- radians 6 classes (45° units) with RD
Sum and variability of incoming solar radiation; Efficiency of weathering activity; Bedrock permafrost; Bedrock hydrology; Snow coverage; Joint orientation				
(Sass, 1998; André, 2002; Dorren and Seijmonsbergen, 2003; Dixon and Thorn, 2005; Gunzburger <i>et al.</i> , 2005; Sass, 2005c,b; Santi <i>et al.</i> , 2009; Wichmann <i>et al.</i> , 2009)				
Morphometry	Profile curvature Surface roughness	DEM (1 m) DEM (1 m)	ordinal metric	3 classes (convex, concave, planar) with RD -
Local stress fields; Sediment storage inside bedrock; Shadowing effects; Local moisture patterns; Snow accumulation; Permafrost distribution				
(Terzaghi, 1962; Gerber and Scheidegger, 1969; Contard and Francou, 1989; Prick, 1997; Sass, 2005c; Matsuoka and Murton, 2008; Shirzadi <i>et al.</i> , 2012)				
Climate and Cryosphere	Annual sum of incoming solar radiation Permafrost probability	DEM (1 m) DEM (1 m) based on PSIM model (Nyenhuis, 2006)	metric metric	- -
Intensity of frost cracking; Thermal weathering processes; Permafrost depth and thermal behaviour; Snow melt and deposition; Bedrock moisture supply; Plant colonisation				
(Hallet <i>et al.</i> , 1991; Matsuoka and Sakai, 1999; Noetzi <i>et al.</i> , 2003; Gruber, 2005; Hall, 2006; Murton <i>et al.</i> , 2006; Hales and Roeting, 2007; Matsuoka and Murton, 2008; Allen <i>et al.</i> , 2009; Sass, 2010; Hall <i>et al.</i> , 2012; Krautblatter <i>et al.</i> , 2013; Viles, 2013; Hall and Thorn, 2014)				
Rock mechanics	Lithology type Joint orientation (in relation to slope)	(Geological map 1:25,000 Swiss Topo, Bearth, 1980) (Geological map 1:25,000 Swiss Topo, Bearth, 1980)	nominal nominal	6 lithological classes with RD 5 joint classes with RD
Rock mass strength; Rock porosity; Moisture distribution; Weathering efficiency; Rockfall type; Rockfall block size; Rockfall frequency (historicity)				
(Selby, 1980, 1982b; Cruden and Hu, 1994; Hoek and Brown, 1997; Cruden and Hu, 1998; Sass, 1998; Wylie and Mah, 2004; Coe and Harp, 2007; Moore <i>et al.</i> , 2009; Fischer <i>et al.</i> , 2012)				
Paraglacial bedrock adjustment	Relative time since deglaciation	Normalised and topography-weighted Euclidean distance to cirque	metric	-
Rock mechanics, i.e. stress, release, formation of sheeting, joints; Weathering historicity; Occurrence, warming and thawing of permafrost; Glacio-isostatic rebound				
(Brunner and Scheidegger, 1973; Augustinus, 1992; Ballantyne, 2002; Cossart, 2008; McColi, 2012; Cossart <i>et al.</i> , 2014; Feuillet <i>et al.</i> , 2014; Leith <i>et al.</i> , 2014)				

faces, where geotechnical information was available, and evaluated in relation to the bedrock's specific slope gradients following Selby (1980) and Cruden and Hu (1994). The following classes are used:

1. Steep joint dipping into the slope ($> 30^\circ$)
2. Moderate joint dipping into the slope (30° -horizontal)
3. Horizontal joint layering
4. Moderate joint dipping out of the slope ($< 30^\circ$)
5. Overdip slope ($> 30^\circ$)

To consider paraglacial rock slope adjustment after the LGM we assume that rockfall activity is highest immediately (or with a short time lag) after deglaciation and is declining (exponentially or linearly) with time (e.g. Curry and Morris, 2004). Due to absent data on glacial retreat stages in the study area, we applied the ergodic reasoning or space-for-time-substitution (Paine, 1985; Pickett, 1989): The time elapsed since deglaciation of rock slopes was approximated by their relative distance from the cirque assuming a gradual retreat of LGM ice from the trough shoulders (Kelly *et al.*, 2004; Ivy-Ochs *et al.*, 2009). For each hanging valley, the relative distance of rockwalls to the cirque was calculated by the normalised Euclidean distance (Fig. 10.3F) ranging from 1 (outlet = deglaciated since ca. 12-18 kyrs BP) to 0 (cirque = recently deglaciated since < 100 yrs). Additionally, to consider ice thinning during deglaciation, which is considerably accelerated on steep relief and retarded on flat terrain, a topographic factor was incorporated in the Euclidian distance calculation. Therefore, the horizontal surface distance to the cirque was weighted by the slope angle at each cell of the 1m DEM. For modelling purposes, all non-metric predictors were converted to metric, except slope gradient, which was recalculated to radians. Non-metric variables were decoded (Table 10.1, Fig. 10.5) using the rockfall densities (RD), as proposed by Bai *et al.* (2010):

$$RD = \frac{\frac{B_i}{A_i}}{\sum_{i=1}^N \frac{B_i}{A_i}} \quad (10.1)$$

where A_i is the area of the i th class of the specific variable, B_i is the total rockfall source area of the i th class, and N is the number of variable classes. Contrary to dummy variables, this approach keeps the original number of predictors. To evaluate the spatial characteristics of rockwalls contributing to talus landforms, the density percentage of the rockfall source areas was calculated for each predictor variable.

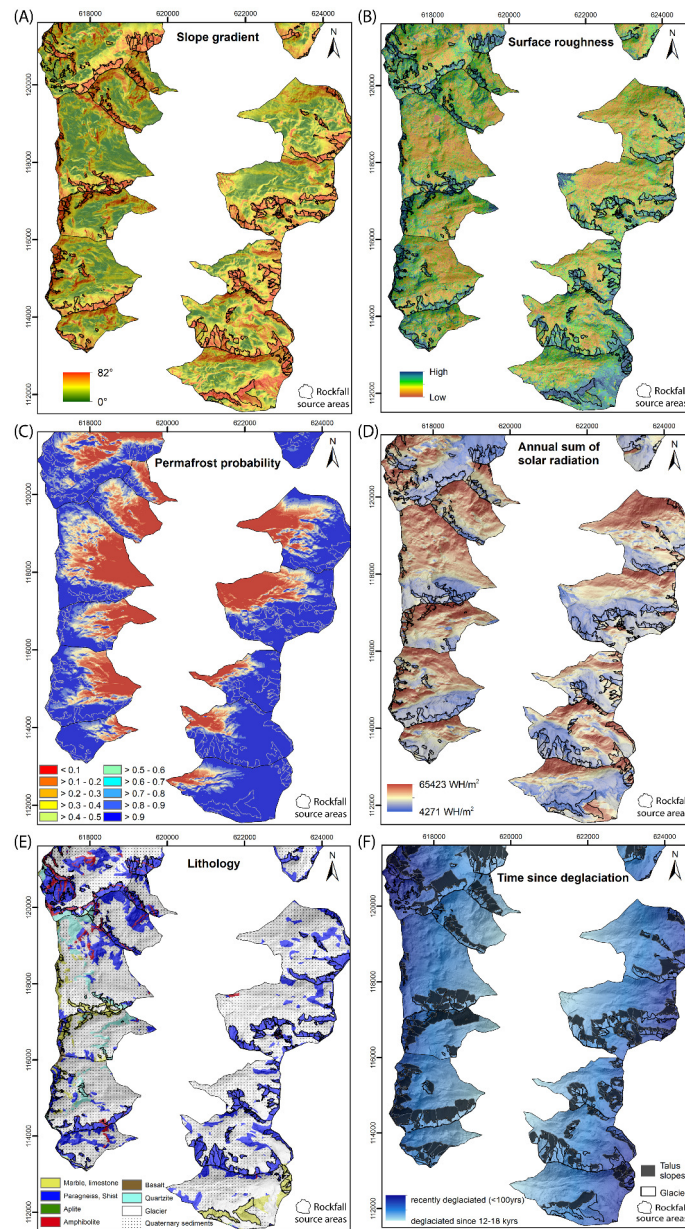


Figure 10.3: Selection of predictor variables representative for topography (A: Slope gradient), morphometry (B: Surface roughness), cryosphere and climate (C: Permafrost probability modelled by Nyenhuis 2006, D: Annual sum of incoming solar radiation), rock mechanical properties (E: Lithology) and paraglacial adjustment (F: Relative time since deglaciation).

10.3.2 Validation methodology

To avoid overfitting during modelling, we split the total dataset into a training data set containing 20% of the pixels, and a validation set containing the remaining 80% of the pixels of the 1 m DTM. Data splitting was done by stratified random sampling to obtain the same distribution of the response variable in both subsets. To validate and compare both model approaches, we estimated

three different validation measures: The accuracy (overall rate of correct classification), the Probability of Detection POD (fraction of observed rockfall source areas that is predicted correctly) and the Probability of False Detection POFD (fraction of observed absent source areas that is predicted falsely as rockfall source areas). The values of validation measures range between 0 and 1, with $POD = 1$ and $POFD = 0$ indicating perfect score (Wilks, 2006). Additionally, a Receiver Operating Characteristic curve (ROC curve) and the specific area under the curve (AUC) serve to visualise the accuracy and prediction power of the model over the complete range. Here, the sensitivity or POD (true-positive fraction) is plotted against the specificity or POFD (false-positive fraction) (Bradley, 1997).

10.3.3 Principal component analysis and logistic regression modelling

When using the multiple logistic regression (LR), our focus is not on the spatial prediction of rockfall probability or susceptibility based on a number of independent predictor variables, but instead on ranking the predictors with respect to their relative importance for the binary dependent response (i.e. presence/absence of rockfall initiation zones). Our algorithm of the principal component logistic regression (PCLR) model was computed using R software packages, based on eight main steps (Fig. 10.4).

Using a logit transformation, the natural log odds (logit) was calculated, being the ratio of the probability of rockfalls (= presence of rockfall source areas) to that of absence. The LR represents a generalised linear regression and can be written as (Atkinson *et al.*, 1998):

$$Y = \text{logit}(p_i) = \log \frac{p_i}{1 - p_i} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n \quad (10.2)$$

where Y is the dependent variable, p_i is the rockfall probability (0 to 1), β_i ($i=1,2,3 \dots n$) is the coefficient of the model, n is the number of independent predictors X_i ($i=1,2,3 \dots n$).

To identify multicollinearity, the Tolerance (TOL) and the Variation Inflation Factors (VIF) were calculated. Variables with $VIF > 2$ and $TOL < 0.4$ might be highly dependent on other predictors (Allison, 2001) and are usually excluded in most regression analyses. The inter-correlation between the permafrost distribution and the time since deglaciation is rather small. This is also supported by their contrasting spatial distribution of these variables shown in Figure 10.3C and 10.3F. In contrast, we detected a moderate inter-correlation between the permafrost probability, slope and aspect (Table 10.2). To solve this problem without a loss of original input data, we used a reduced set of uncorrelated PCs represent-

ing linear combinations of the ten original standardised data with a maximum possible variance (Escabias *et al.*, 2005; Aguilera *et al.*, 2006).

To select only those principal components (PCs) with statistically significant contribution to rockfall activity a forward stepwise method was applied using maximum likelihood ratio ($-2\ln L$) tests together with the Akaike Information Criterion (AIC) value. Here, PCs were selected based on their predictive ability and not on the largest variance (Aguilera *et al.*, 2006). Starting with the null-model (intercept only), PCs were successively added until further additions did not result in a lower $-2\ln L$ and AIC value. Using the Wald X2 statistics, the predictors' coefficients were estimated to be statistically significant, if the tested null hypothesis (H_0 : the estimated coefficient is 0) could be rejected at a $p = 0.001$ significance level (Kleinbaum *et al.*, 1998). The association between the predicted probability and the observed responses (goodness-of-fit) was tested over the training dataset using the X2 value (based on $-2\ln L$) of the Hosmer-Lemeshow test, the $-2\ln L$ and the AIC value (Hosmer and Lemeshow, 2005).

For validation of the fitted logistic regression model we used a contingency table using the validation data set based on a cut-off value of 0.509. This threshold was determined by finding the best trade-off between sensitivity (Probability of Detection) and specificity (1- Probability of False Detection) of the predicted probabilities, i.e. the trade-off between failing to detect rockfall against the costs of raising false alarms. The strength of association between response and each predictor was determined by means of the Odds ratios (OR) of the varimax-rotated PCs (Kaiser, 1958) with respect to the loadings of the original standardised predictors. The OR is the exponential of the regression coefficient (e^{β_i}) associated with a one-unit increase in X_i . If a coefficient is positive, then OR is > 1 and thus, rockfalls are more likely to occur. In turn, the likelihood of rockfalls decreases with negative coefficients and $OR < 1$. In case of no or weak causal link between predictors and response, OR is ≈ 1 (Atkinson *et al.*, 1998).

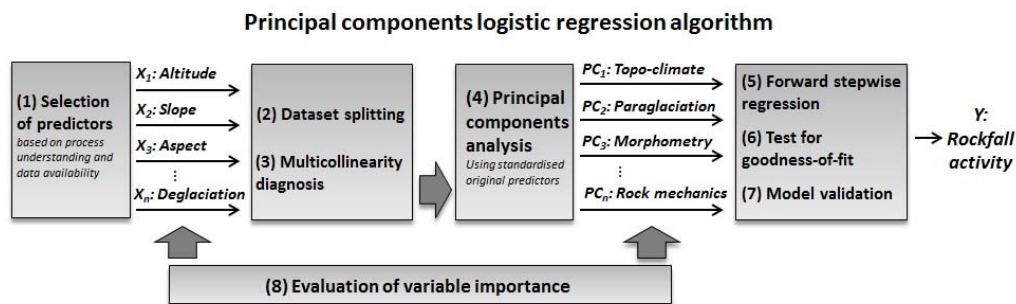


Figure 10.4: Modelling approach of the multiple logistic regression using uncorrelated principal components (PC) instead of the original, intercorrelated predictor variables X .

Table 10.2: *Diagnostics statistics of multicollinearity between independent predictors using the Tolerance (TOL) and the Variation Inflation Factors (VIF). VIF of > 2 and TOL of < 0.4 point to variable intercorrelation.*

Predictor variables	VIF	TOL
Elevation	1.95	0.51
Slope	4.96	0.20
Aspect	4.24	0.31
Curvature	1.01	0.99
Roughness	1.42	0.71
Solar Radiation sum	2.96	0.30
Permafrost probability	6.16	0.16
Lithology	1.17	0.85
Joints	1.10	0.90
Deglaciation	1.94	0.51

10.3.4 Random forest model

Machine learning algorithms such as the random forest algorithm of Breiman (2001a) are known as being able to deal with complex interacting as well as highly correlated predictor variables. The RF model is based on the concept of classification trees. Tree-based models consist of a series of nested decision rules for the predictors that determine the response. Random forest repeatedly builds trees from random samples of the training data. Each tree is treated as a separate model of the ensemble. The majority class of all trees is taken as final estimate of the model. To overcome correlation between trees, a certain number of predictors (commonly abbreviated as “MTRY” in the random forest literature) are randomly selected at each split of a tree. The best predictor from the random subset is used at the respective split to partition the data. MTRY is a parameter, which must be adapted to the respective model data. We used the R implementation of the random forest algorithm (Liaw and Wiener, 2002) to classify the rockfall source areas based on the set of predictor variables. Following the suggestion of Kuhn and Johnson (2013a), MTRY was tuned between 2 and the number of predictors. The training was performed using a stratified 10 fold cross-validation. Therefore, models were fitted by repeatedly leaving one of the folds out. The performance of each model was determined by predicting on the respective withheld fold. The ROC from the withheld iterations was averaged to the overall performance for the respective value of MTRY. The model resulting from the best set of tuning parameters was used as the final model for prediction

on the test data. Variable importance was further calculated as described in Liaw and Wiener (2002), using the mean decrease in GINI index. The GINI index is a criterion of impurity of a node in the RF model (Breiman, 2001a). At every split of a node, one of the randomly chosen variables is used to form the split, which results in a decrease in the GINI index. Thus, the mean decrease in GINI index over all trees in the forests indicates how much the given predictor variable contributes to the impurity of nodes in the model. The higher the decrease in GINI index, the higher the purity of the final RF model, and thus the more important is the predictor variable.

10.4 Results

10.4.1 Spatial characteristics of rockfall source areas (rockfall density statistics)

The rockfall density percentage (RD %) of active initiation zones with respect to each predictor variable is shown in Figure 10.5. Around 11% of the total bedrock area in the hanging valleys represents active rockfall source areas contributing to active talus landforms. The rockfall source zones predominate on NNE-NNW-exposed slopes (41% RD, Fig. 10.5C) with slope gradients $> 40^\circ$ (93% RD, Fig. 10.5B). The dominance of north-facing initiation zones is independent of the morphometric configuration of the study area, given the relatively uniform orientation of the total bedrock area across the hanging valleys. Half of the active rockfall source areas receive very low incoming solar radiation of less than 14300 WH/m² per year (Fig. 10.5D). The majority of rockfalls occur in the upper half of the hanging valleys (Fig. 10.5J), although bedrock slopes dominate in the middle basin at 2500-2800 m a.s.l. About 74% of the rockfall initiation zones are primarily concentrated within an elevation belt at 2900-3300 m a.s.l. (Fig. 10.5A). This is consistent to the lower boundary of the modelled permafrost distribution. Particularly, 57% of the unstable rockwalls correlate with a 90-100% permafrost probability (Fig. 10.5E). Considering the micro-topography, active rockfall source areas are equally concave and convex and 70% are characterised by very rough bedrock surfaces (Fig. 10.5F, G). Despite the dominance of paragneiss and micashist in the study area, only 14% of the active rock slopes can be found there (Fig. 10.5H). Instead, failures mostly occur in amphibolite (25% RD) and limestone (20% RD), contrasting to their small relative catchment areas (0.2% amphibolite, 12% limestone). Finally, half of the rockfall initiation zones lie at anacinal slopes with moderately and steeply in-dipping joints, while 36% are linked to cataclinal slopes with out-dipping joints (Fig. 10.5I).

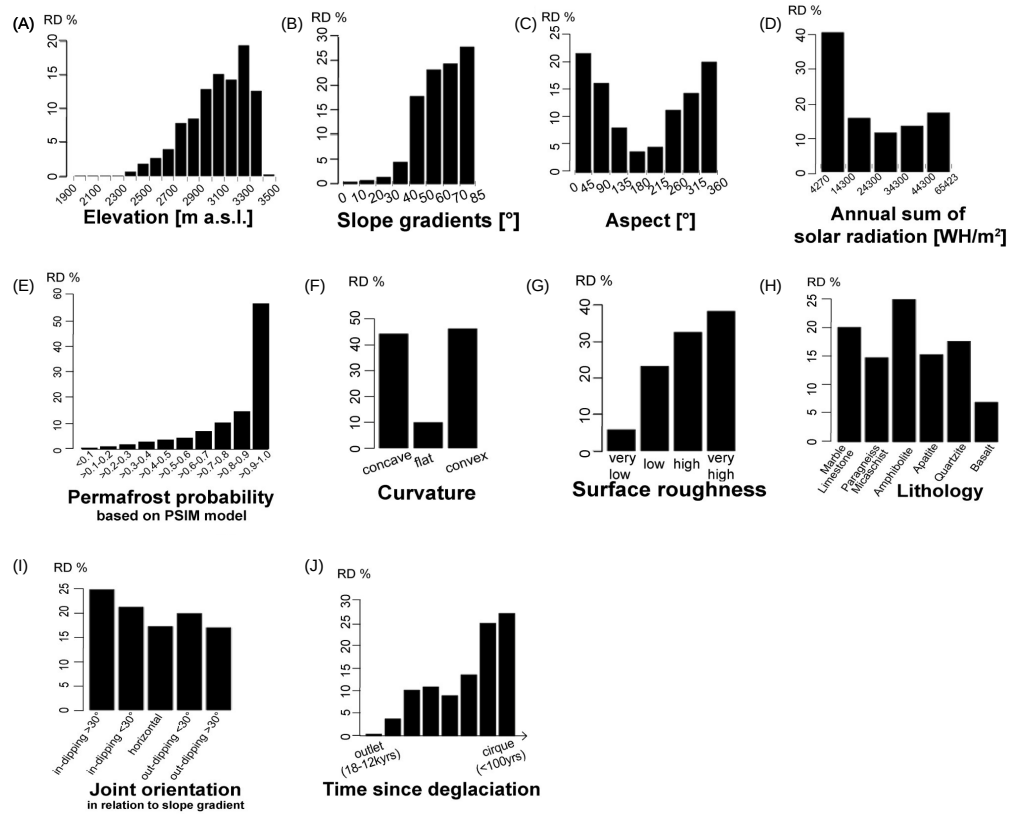


Figure 10.5: Rockfall densities (RD in %) of predictor variables including topographic (A-C), climatic and cryospheric (D-E), morphometric (F-G), rock mechanical (H-I) and paraglacial (J) properties.

10.4.2 Principal components and geomorphic meaning

The first eight varimax-rotated principal components account for around 98% of the total variability of the original, standardised predictor dataset (Table 10.3). Most of the variance is explained by PC1 (20%) and PC3 (18%), respectively, while the contribution of PC9 and PC10 is < 1%. PC1, PC3 and PC9 represent topo-climatic characteristics. The first component is strongly related to steep slope gradients (loading of 0.95) and a high probability of permafrost (loading of 0.85). PC3 is highly associated with convex S-exposed slopes and high annual solar radiation sums, but negatively correlated with permafrost occurrence. PC2 denotes paraglacial adjustment given the high loading of the time since deglaciation. Likewise, PC8 is strongly correlated with elevation and, to minor extent, with the time since deglaciation. The bedrock morphometry is represented both in PC5 and PC7, with a high positive loading of curvature (0.98 in PC5) and roughness (0.93 in PC7). PC4 describes rock mechanical properties, given the very high loading of in-dipping joints. The type of lithology is characterised by PC6, strongly associated with amphibolite and limestone rocks. Finally, PC9 summarises N-facing, low-radiation slopes, whereas PC10 stands for permafrost

occurrence; but the predictors' loadings are very low (< 0.26), respectively.

10.4.3 PCLR model and importance of PCs

In the stepwise logistic regression procedure (stepwise model I), PC4, PC9 and PC10 were excluded, since their elimination lead to a significant reduction ($p=0.001$) of $-2\ln L$ and AIC. This decrease of significance was greater compared to the full model including these three PCs (Table 10.4). To evaluate the goodness-of-fit when integrating rock mechanical properties (PC4), we additionally tested a stepwise model II.

Both stepwise models seem to perform very well as a significant decrease of $-2\ln L$ and AIC was obtained in each stepwise procedure in comparison with the full model (containing all PCs) and the null-model with intercept only. Additionally, the Wald X2 statistics indicate an acceptable model performance of the stepwise models at a significance larger than 0.001. However, the incorporation of PC4 does not result in a distinctly better model fitting, indicating that the joint orientation might not be necessarily causative for the spatial rockfall pattern. This is also demonstrated by the AUC value of stepwise model II, which is 1% lower compared to stepwise model I. Consequently, we decided to exclude PC4 from the final regression. Therefore, the best LR model for predicting the spatial pattern of rockfalls includes seven PCs (based on Eq 2):

$$\begin{aligned} \text{Logit}(p) = \log(p/1 - p) = & 0.23 + 1.08PC1(\text{"Topo - climate"}) + \\ & 0.26PC2(\text{"Paraglacialadjustment"}) - 0.89PC3(\text{"Topo - climate"}) + \\ & 0.11PC5(\text{"Bedrockmorphometry"}) + 0.29PC6(\text{"Lithology"}) + \\ & 0.28PC7(\text{"Bedrockmorphometry"}) + 0.38PC8(\text{"Paraglacialadjustment"}) \end{aligned} \quad (10.3)$$

The Wald X2 statistics reveal that all coefficients are statistically significant at $p < 0.001$. Comparing the ORs of the PCs (Table 10.4), PC1 was detected to be most strongly associated with the spatial rockfall activity, directly followed by PC3. One-unit increase of these topo-climatic PCs therefore leads to a 2.94 times higher or 0.41 lower rockfall likelihood, respectively. The third most important control is PC8, increasing the odds of rockfalls by 1.47 times. The ORs of PC2, PC6 and PC7 range between 1.34 and 1.29 indicating a positive, but comparatively moderate impact on the spatial rockfall pattern. Finally, the OR of PC5 is relative close to 1, indicating a weak power to explain the response variable.

Table 10.3: *Varimax-rotated principal components of original standardised predictor variables. Geomorphic meaning of PCs is based on strength of factor loading of original predictors. PC1, PC2, PC3, PC5, PC6 and PC7 were selected to be significantly causative with the response variable in a stepwise regression modelling based on the -2LnL and AIC value.*

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
Geomorphic meaning	Topo-climate	Paraglacial adjustment	Topo-climate	Rock mechanics	Bedrock morphology	Lithology	Bedrock morphology	Paraglacial adjustment	Topo-climate	Permafrost
	Steep slopes, high permafrost probability	Very short time since deglaciation, high elevation	S-exp. slopes, high solar radiation, medium permafrost probability	In-dipping (>30°) joints	Convex bedrock profiles	Amphibolite and limestone types	Very rough bedrock surface	High elevation, Short-medium time since deglaciation	N-exp. slopes with medium solar radiation	Low to medium permafrost probability
Elevation	0.14	0.36	0.12	0.06	0.02	0.17	0.07	0.91	0.01	0.00
Slope	0.95	-0.01	0.18	-0.03	0.10	0.12	0.20	0.12	-0.02	0.15
Aspect	0.25	-0.04	-0.76	0.03	0.00	0.04	0.02	0.05	0.23	-0.03
Curvature	0.10	0.02	0.72	-0.01	0.98	0.03	0.14	0.02	0.00	0.00
Roughness	0.30	0.03	0.07	-0.04	0.17	0.08	0.93	0.06	0.00	-0.01
Solar radiation sum	0.25	0.03	0.95	-0.03	-0.03	-0.06	-0.06	0.07	0.23	-0.02
Permafrost probability	0.85	0.07	-0.43	0.05	-0.08	-0.07	-0.20	0.00	-0.03	0.26
Lithology	0.14	0.06	0.08	-0.10	0.03	0.97	0.07	0.14	0.00	0.00
Joints	-0.06	0.10	0.05	0.99	-0.01	-0.09	-0.04	0.05	0.00	0.00
Time since deglaciation	0.06	0.94	0.05	0.11	0.01	0.06	0.02	0.31	0.00	0.00
Eigenvalues	2.98	1.03	1.03	1.01	1.01	1.01	0.98	0.96	0.11	0.09
Explained Variance	0.20	0.10	0.18	0.10	0.10	0.10	0.10	0.10	0.01	0.01
Cumulative Variance	0.20	0.30	0.48	0.58	0.68	0.78	0.88	0.98	0.99	1

Table 10.4: *Test for goodness-of-fit using the -2log likelihood ratio tests (-2LnL), Akaike Information Criterion (AIC), X2 and area under the ROC curve (AUC).*

Model	-2LnL	AIC	X2	Pr(> X ²)	AUC
Null model (intercept only)		25811			0.50
Full model (with all PCs)	19258	19275	19257	< 0.0001	0.80
Stepwise model I PC1. PC2. PC3. PC5. PC6. PC7. PC8	19247	19265	19247	< 0.0001	0.83
Stepwise model II PC1. PC2. PC3. PC4. PC5. PC6. PC7. PC8	19247	19266	19247	< 0.0001	0.82

Table 10.5: *Coefficient statistics: β = logistic regression coefficient of PC. S.E. (β) = standard error on β . Wald X2, P (>| X2|), odds ratio = $e^{\beta i}$ with a S.E. of ± 1.96 and a 95% confidence interval (C.I.). The variable importance was ranked from I (most important) to VII (less important) based on the odds ratio.*

Variable	β	S.E. (β)	Wald X2	Pr(> X2)	OR	95% C.I.	Variable importance – ranking (I-VII)
(Intercept)	0.23	0.02	13.00	<0.0001			
PC1	1.08	0.02	50.40	<0.0001	2.94	2.82 - 3.07	I
PC2	0.26	0.02	14.11	<0.0001	1.29	1.25 - 1.34	VI
PC3	-0.89	0.02	-47.79	<0.0001	0.41	0.40 - 0.43	II
PC5	0.11	0.02	6.09	<0.0001	1.12	1.08 - 1.16	VII
PC6	0.29	0.02	15.57	<0.0001	1.34	1.29 - 1.39	IV
PC7	0.28	0.02	13.97	<0.0001	1.32	1.27 - 1.38	V
PC8	0.38	0.02	21.38	<0.0001	1.47	1.42 - 1.52	III

10.4.4 Random forest model and variable importance

The tuning of the random forest model revealed an optimal MTRY value of 8. The variable importance indicated by the GINI decrease denotes the regional distribution of permafrost as the most important predictor (Fig. 10.6). Further, the time since deglaciation and the elevation are shown to have considerable influence on the classification result. Furthermore, the sum of solar radiation, the slope gradient, bedrock roughness and slope aspect are linked to a medium relative variable importance. The two rock mechanical predictors including joint orientation and lithology, followed by curvature, were ranked as least important for the spatial distribution of rockfall source areas.

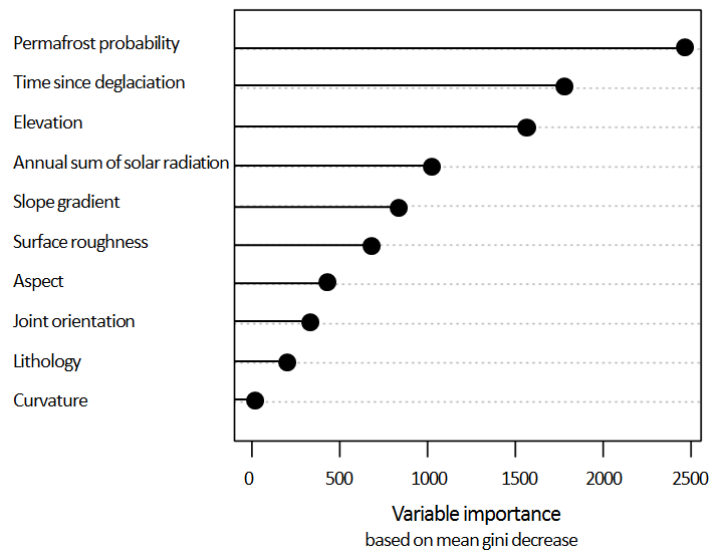


Figure 10.6: Variable importance quantified by means of the random forest mode using the mean decrease in GINI index. The higher the decrease in GINI index, the more important is the variable for the spatial activity of rockfalls.

10.4.5 Validation and evaluation of model performances

To validate the two models with respect to each other we estimated the same validation measures using the same validation data set (Table 10.6). For the PCLR, the overall rate of correct classification is estimated as 75.52%, which is considered to be very acceptable. While 73.71% of the rockfall initiation zones are correctly predicted (POD), the POFD is 22.48%. A comparatively better accuracy is estimated for the RF as 88.40% of the positive and negative observations are correctly predicted. Likewise, the RF results with 88.40% in a comparatively higher POD and only 11.60% is false alarm (POFD). The validation by means of the ROC curves (Fig. 10.7) reveals for both models a very good performance given their high AUC values > 0.80 (Swets, 1988). However, the AUC value of the RF model (AUC = 0.95) is slightly higher compared to the PCLR (AUC =

0.83), reflecting a slightly better performance of the machine learning algorithm.

Table 10.6: Contingency table for (A) the principal component logistic regression model and (B) the random forest model. Three verification measures were estimated: Accuracy, Probability of Detection (POD) and Probability of False Detection (POFD).

(A) PCLR		Observed		
		Yes (1)	No (0)	Total
Predicted	Yes (1)	33.71	12.11	45.82
	No (0)	12.36	41.76	54.12
	Total	46.07	53.87	99.94
Accuracy = $(33.71 + 41.76) / 99.94$; POD = $33.71/46.07$; POFD = $41.76/53.87$				
(B) RF		Observed		
		Yes (1)	No (0)	Total
Predicted	Yes (1)	40.78	6.25	47.03
	No (0)	5.35	47.62	52.97
	Total	46.13	53.87	100
Accuracy = $(40.78 + 47.62) / 100$; POD = $40.78/46.13$; POFD = $47.62/53.87$				

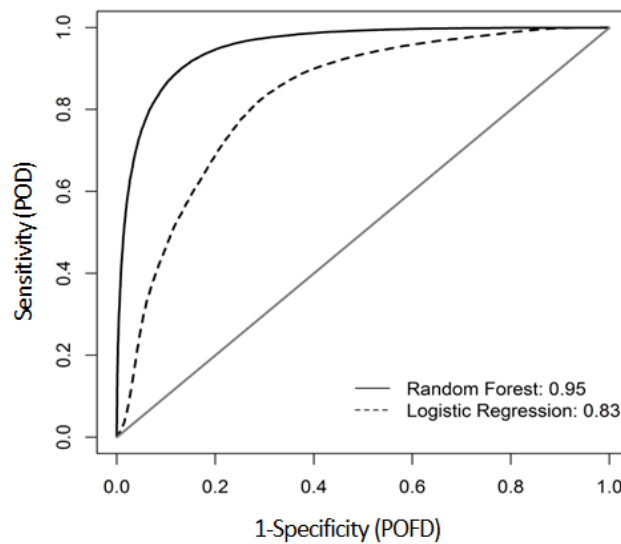


Figure 10.7: Receiver Operating Characteristic (ROC) curves and values of the corresponding areas under the curve (AUC) of the random forest and logistic regression model.

10.5 Discussion

10.5.1 Evaluation of the methodological approach

Given the large areas under the ROC curves of 0.83 and 0.95 (Fig. 10.7), respectively, the logistic regression and the random forest classification have proven to be very effective techniques to explain the spatial rockfall pattern in the Turtmann Valley. Nevertheless, there are some systemic uncertainties and limitations underlying our approach. Admittedly, using an inventory of presently vegetation-free talus slopes as proxy to infer potential rockfall initiation zones and their controlling factors is only valid under the assumption that recent talus slope deposition very likely resulted from the same conditions, which have been causative in the past (Varnes, 1984; Carrara *et al.*, 1999). This abductive reasoning might be true regarding large-scale, relatively static tectonic, lithological and topographic settings; but obviously, weathering history (Viles, 2013), climatic and cryospheric conditions (Gruber and Haeberli, 2007), bedrock morphometry and rock mechanical properties (Verleysdonk *et al.*, 2011) highly evolved since the LGM (Fig. 10.1), all integrated in the evolution of talus slopes. However, the changing relative contribution of rockfall controls over time remains difficult to quantify (Church, 1996; Guzzetti *et al.*, 1999) and we must argue actualistically and often narratively. Similarly, it is important to note that the spatial occurrence of rockfalls crucially depends on the point of observation time. Rock slopes, which we identified as being recently inactive, were likely affected by rockfalls in previous times, e.g. when the permafrost boundary was lower than today. To overcome this, our approach implements an ergodic reasoning to use the spatial pattern of active/inactive rockwalls for the temporal shift in rockfall activity since LGM. Further, the detection of rockfall source areas using the hydrological GIS approach is linked to some restrictions, as falling of rock fragments from steep slopes cannot be simply equated by the water-driven flow paths. However, Duarte and Marquinez (2002) found a good agreement between a similar automatic method and a manual mapping of detachment zones, implying that our approach might be applicable at larger scales. Alternatively, it would have been possible to use the approach proposed by Heckmann *et al.* (2016), who delineated potential rockfall source areas by combining their slope angle distribution derived from a high resolution DEM (Loye *et al.*, 2009) with a field-based geomorphological map.

Additionally, the pre-selection of predictors certainly depends on data availability, raster resolution and technical restrictions typical for spatial modelling (Carrara *et al.*, 1999). For instance, we compensated the missing data on timing of deglaciation by the relative Euclidean distance to the cirques. This ergodic reasoning has been successfully applied for studying hillslope evolution (Carson

and Petley, 1970; Brunsden and Kesel, 1973; Obanawa *et al.*, 2009), river system changes (Fryirs *et al.*, 2012) and vegetation succession (Cammeraat *et al.*, 2005). Further, the calculation of solar radiation by means of the hemispherical viewshed GIS algorithm (Fu and Rich, 2002) might be rather idealised, as a disturbed solar transmission through atmosphere and topographic shadowing effects are not taken into account (Allen *et al.*, 2009). The permafrost validation of the PSIM by means of a local rock glacier inventory revealed a good agreement of ca. 87% , which was even larger when using classical methods like ROGMOD or PERMAP (Nyenhuis *et al.*, 2005). However, the rock glacier based modelling probably underestimates the subsurface temperatures of steep rock slopes and the lower limit of permafrost is probably higher than in more gentle terrain (Magnin *et al.*, 2015), given the specific conductive and advective thermal fields inside the anisotropic rock mass (Gruber, 2005; Noetzli and Gruber, 2009). Likewise, despite their recognised relevance for slope instability, bedrock moisture (Sass, 2004, 2005a), snow cover (Draebing *et al.*, 2014), biological influences (Chen *et al.*, 2000; Hall *et al.*, 2005) or discontinuities at a cm-dm scale resulting from thermal fatigue (Hall and Thorn, 2014) and stress release (Augustinus, 1995) cannot be portrayed at the regional scale. Besides methodical limitations, variable selection also arises from a prior systemic knowledge on rockfall mechanisms, leading to the risk filtering out any factors being underestimated in their efficiency or unknown so far (Hall, 2006).

Finally, with respect to our sampling strategy (20%-80% sample splitting on pixel basis), it is to note that the cross validation error metrics might tend to be somewhat overoptimistic. Assuming that the neighbouring pixels in the vicinity of the rockfall initiation zones have similar properties, the samples are not completely independent. To overcome this issue, it might be an option to consider stricter cross validation methods in upcoming studies. However, since the main aim of this study is not to predict the spatial distribution of rockfall activity, but to quantify the relative importance of various controlling factors, the cross validation does not change the general outcome of this study.

Taking into account these uncertainties, our proposed modelling approach is promising to evaluate the regional-scale causality between rockfalls and potential key drivers. Instead of avoiding multicollinearity of data, which is often ambiguous or intuition driven by the researcher's pre-existing hypotheses, far more importance should be given to its evaluation (cf. Graham, 2003), as multicollinearity is typical in complex non-linear geomorphic systems. Using PCs as covariates in the logistic regression allowed keeping all original predictors in the regression and preventing a substantial loss of explanatory power, as reflected very well by the high POF (75%) and low POFD (22%) of our final LR model (Table 10.6). Furthermore, while most (landslide susceptibility) studies treat the relative predictor importance as "by-product" of the logistic regression modelling,

our PCLR algorithm (Fig. 10.4) has proven to be a powerful tool to identify characteristic synergetic combinations of different rockfall controls with respect to the strength and direction of their association. The PCs' varying predictor loadings (Table 10.3) efficiently contribute to a better understanding of whether rockfalls are more affected by rock mechanical, paraglacial or topo-climatic characteristics. Admittedly, the geomorphic interpretation of the association between response and predictors is less straightforward and needs a profound systemic knowledge.

To complement the PCLR model and to compensate some constraints in interpretation we applied a second modelling, the decision-tree based random forest algorithm, which is one of the most precise machine learning algorithms. To the authors' knowledge, we are the first to use the RF model for rockfall analyses. Compared to the LR, all validation indices (Table 10.6, Fig. 10.7) imply an overall better performance and higher accuracy of the RF when explaining the spatial rockfall pattern in our study area. With respect to the computation time, it is to note that logistic regression training and prediction took 17 seconds on one core while the RF modelling took 15 minutes even on 4 cores. However, the computational complexity of the tree-based classification is compensated by the relatively rapid quantification and interpretation of the predictors' relative importance (Fig. 10.8). While the decision-tree based model works more like a "black-box", the PCLR requires a conceptual understanding of the geomorphic system to name and identify the principle components. The RF approach is also highly attractive since it accepts the multicollinearity in our dataset without using PCs, and allows mixing of categorical and metric variables without decoding (Catani *et al.*, 2013). Our study demonstrates therefore the great potential of the RF algorithm for future applications in rockfall research.

Given the coherent results of both models, we conclude that the classical logistic regression can perform comparably with the novel, but often time-consuming and technically challenging machine learning approaches (*sensu* Brenning, 2005; Vorpahl *et al.*, 2012). However, to achieve reliable results we favour the application of multiple models relying on different degree of systemic knowledge on the geomorphic system. Here, by combining a random forest black-box approach with a strongly knowledge- and theory-based logistic regression of principal components we hope to avoid filtering out anything that we do not expect or do not want to see (Hall, 2006) as well as to argue purely driven by statistics.

10.5.2 Regional-scale controls on rockfall activity

10.5.2.1 The predisposing effect of rock mechanical characteristics

The influence of structural and lithological properties on the regional-scale rockfall activity is poorly studied and largely underestimated, hampered by the

prevailing idea of environmental forcing or due to the methodical challenge of spatial extrapolation of geotechnical data. However, the growing availability of high-resolution DEMs and digital geological maps as well as the improvements in GIS and image analyses technology promise considerable advances for regionalisation of structural bedrock characteristics, even in steep mountain terrain (Jaboyedoff *et al.*, 2007; Günther *et al.*, 2012; Matasci *et al.*, 2015). In the hanging valleys of the Turtmann Valley, the lithological and structural control is reflected very well in our results. Despite the dominance of metamorphic rocks, amphibolite and limestone are found to be most sensitive to rockfalls (Fig. 10.5H). These findings may seem surprising as their high compressive strength (Selby, 1980; Wyllie and Mah, 2004) and low joint density (Sass, 2005b) would suggest a relatively low erodibility. In the Cantabria Mountains, e.g., Duarte and Marquinez (2002) showed that siliciclastic rocks with a high density of open joints are more affected by rockfalls than limestone of lower joint densities. However, as stated by Fischer (2010), it is often exactly the small-scale transitions between different lithological units that effectively promote failure due to the contrasting hydraulic regimes and stress-strength conditions (Evans and Hungr, 1993).

We furthermore found a relatively strong relationship between rockfall scars and in-dipping joints (Fig. 10.5H,I). This also contrasts to common rock mechanical studies (e.g. Selby (1982b); Moore *et al.* (2009), identifying cataclinal slopes with surface-parallel joints as more sensitive for instability due to the pre-existence of sliding planes (Cruden and Hu, 1998). However, failure mechanisms in our study site differ from those in other studies. Resulting from the high compressive strength (e.g. amphibolite) and in-dipping, widely spaced joints, detachment mostly occurs through toppling processes of large blocks, as being reflected in the large and cubic blocks on talus slopes (Fig. 10.8A). In contrast, highly weathered paragneiss coupled with cataclinal bedding and high joint densities is linked with surficial flaking off, which is only effective enough to accumulate block slopes of platy, small-size debris, as observable on the southern valley flanks (Fig. 10.8B).

Compared to topo-climatic and paraglacial variables, the relative contribution of rock mechanical properties to the regional-scale rockfall pattern in the Turtmann Valley was minor. Although the existence of amphibolite and limestone rocks can increase the susceptibility for bedrock failure by a factor of 1.34 times (Table 10.5), this is low relative to the other predictors' impact in the PCLR model. Similar findings were obtained by the RF model, relating joint orientation, lithology and curvature (leading to local stress fields) to the lowest explanatory power (Fig. 10.6). However, although our modelling might reveal that rock mechanical properties are currently not the most effective rockfall driver at regional scale, lithology and joint orientation certainly have a major predisposing control, as also shown for catchments in the Eastern Italian Alps (Frattini *et al.*,

2008) and Norway (Böhme *et al.*, 2014). Below, we will further evaluate how the relative contribution of paraglacial and environmental factors is significantly preconditioned by the rockwalls' mechanical properties.

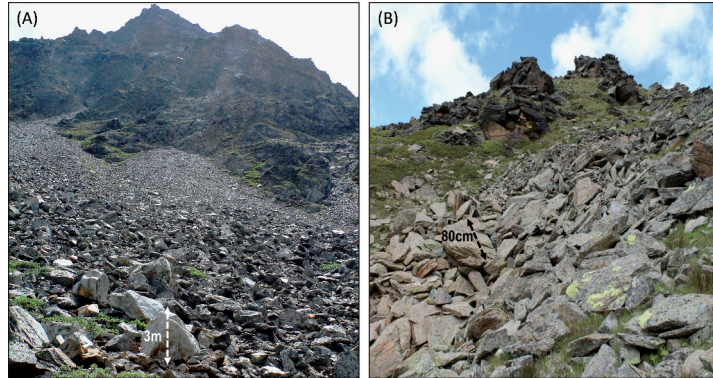


Figure 10.8: Typical examples for active talus slope deposition along north-exposed rockwalls (A) and mainly inactive, vegetated rockfall deposits along south-exposed rock slopes (mainly outcrops) (B). At north-facing, probably permafrost-affected rockwalls, high compressive strength of amphibolite (and limestone) and in-dipping joint bedding might temporary increase the rock slope stability. However, if rock breakdown e.g. by frost cracking occurs, toppling processes of large block sizes (10-104m³) are supposed to be the major failure mechanism, reflected by the large block sizes (3m in b-axis) and cubic forms of the talus slope material. At sun-faced slopes, low compressive strength of paragneiss rocks and cataclinal bedding in combination with frequent thermal, but dry cycles mostly result in flaking off of near-surface bedrock (cm-mm). As results, talus slopes are mainly made of small-sized (80cm in b-axis) and platy rock material. The lichen cover might reflect the inactivity of rockwall today, however, it is also to note that it might be caused by the warmer climate conditions at the south-exposed slopes.

10.5.2.2 Paraglacial adjustment processes as system inherent controls

The obvious dominance of rock slope failures in deglaciated mountain geosystems is often causally linked to the disappearance of glaciers (Bovis, 1990; Cossart, 2008; Ballantyne and Stone, 2013). Former glaciation and subsequent deglaciation is assumed to have conditioned mountain rock slopes, rapidly adjusting to a nonglacial equilibrium through enhanced rockfall events (Church and Ryder, 1972; Ballantyne, 2002). However, compared to non-glacial factors, the specific role of paraglacial bedrock adjustment for both local- and regional-scale rockfall activity is still subject to major uncertainties (McColl, 2012). Our analyses give strong evidence for a paraglacial forcing on rockwalls in the Turtmann Valley. In the RF model, the time since deglaciation is the second most important predictor (Fig. 10.6). A similar explanatory power was detected in the PCLR model, where PC8 (paraglacial adjustment) is the covariate with the third highest odds ratio (Table 10.5). With increasing time elapsed since deglaciation and increasing elevation, failure susceptibility of bedrock can increase up to 1.29 (PC2) and

1.47 times (PC8), which is comparable to findings from Great Britain (Hinchliffe and Ballantyne, 1999; Curry and Morris, 2004) and Svalbard (André, 1997). The paraglacial bedrock adjustment is reflected very well in the characteristic rockfall pattern showing a significant overrepresentation in the upper basins (Fig. 10.5J), where rock slopes have been deglaciated during the last <100 years. In turn, rock faces near the trough shoulders, where deglaciation started earlier ca. 18-12 kyrs ago, might have already worked off the paraglacial signal, given the recent rockfall inactivity there.

With respect to the timing of paraglacial rock slope adjustment, the regional-scale activity of rockfalls, shown in Figure 10.5J, appears to confirm the exhaustion model proposed by Cruden and Hu (1998), assuming a rockfall peak immediately after deglaciation and a fast decline (Fig. 10.9, curve A). While most large-scale paraglacial studies ignore or avoid possible rock mechanical influences (Ballantyne and Stone, 2013; Ballantyne *et al.*, 2014), our data reveal that the timing and intensity of paraglacial rockfall events might depend on the specific tectonic and geological settings, i.e. on the general SW-dipping of the metamorphic rocks. At a slope scale, the comparatively lower rockfall susceptibility of recently deglaciated south-facing rockwalls (Fig. 10.5C) could imply that cataclinal slopes with low compressive strength (e.g. weathered paragneiss) favour a very rapid paraglacial response due to the prompt initiation of pre-existing shear planes. As soon as the weathered rocks are eroded, the slopes may immediately adjust to a non-glacial strength equilibrium (Fig. 10.9, curve C). In contrast, it is reasonable to assume that along north-exposed rockwalls, whose stability is temporally increased due to high internal strength of amphibolite and anacinal bedding opposite to the general SW-oriented tectonic rock layering, rockfalls may start delayed, rather weaker (Fig. 10.9, curve B) after deglaciation. As consequence, paraglacial rockfall activity likely diminished more linearly, supporting Cruden and Hu's (Cruden and Hu, 1998) idea of a constant frequency (steady state) model. Today, the process interplay responsible for paraglacial adjustment is still disputed and shows a significant time- and scale-dependence (McColl, 2012). At the scale of individual rock slopes, our modelling analyses do not allow further conclusions on whether glacial unloading of rockwalls and post-glacial debuttreasing resulted in relaxation of internal stresses (Evans and Clague, 1994; Augustinus, 1995) and in propagation of stress-release joints parallel to the former glacier contact zone (Lewis, 1954; Bovis, 1990; Hencher *et al.*, 2011). Yet at larger scales, the obvious higher rockfall activity detected at nearly vertical slopes ($> 60^\circ$, Fig. 10.5B) appears to support the hypothesis that glacial over-steepening relative to the internal rock mass strength is a major paraglacial driver in our study area (Augustinus, 1992; Leith *et al.*, 2014). Furthermore, a glacio-isostatic rebound might be plausible for landslides clustering over large geographic areas, as calculated by Feuillet *et al.* (2014) and Cossart *et al.* (2014)

for northern Iceland, but can likely be precluded here. A post-glacial uplift of hanging valleys is probably not variable enough to produce the heterogeneous rockfall pattern or is likely overlain by the large-scale uplift of the Rhone valley (0.2-1.6 mm/a Gudmundsson, 1994; Schlunegger and Hinderer, 2001). Instead, paraglacial rockwall adjustment appears to be strongly related to warming and thawing of permafrost, given the obvious dominance of rockfall scars within a low-solar elevation belt between 2900-3300 m (Fig. 10.5A, D), consistent with the regional permafrost distribution (Nyenhuis, 2006). In findings from the Southern Alps of New Zealand (Davies *et al.*, 2001; Hales and Roering, 2007; Allen *et al.*, 2009), glacier retreat during LGM was probably accompanied by a gradual rise of the lower permafrost level and, concurrently, of the frost cracking window; even though this may not be necessarily true for S-facing slopes. According to calculations by Hales and Roering (2009), a temperature drop of 4 °C, consistent with possible climatic conditions during LGM in the New Zealand Alps, would lead to 500 m lowering of the zone, where warming and thawing of permafrost effectively promote changes in internal rock shear strength and water pressure (Krautblatter *et al.*, 2013). This could explain the former rockfall activity of the vegetated hillslopes near the trough shoulders (Fig 10.5J, 10.8B). Therefore, when evaluating the causality between climate change, permafrost degradation and rockfalls, the rock mass' memory effect must be considered, as slope instability is often a delayed response to paraglacial forcing on rock mechanical and thermal systems (Krautblatter and Moore, 2014).

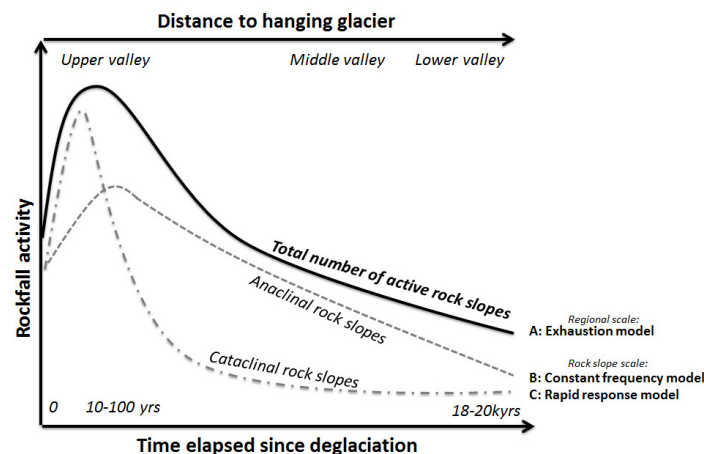


Figure 10.9: Possible models for the timing of paraglacial rockfall activity in the Turtmann Valley referring to Ballantyne and Stone (2013), p. 151. Depending on the spatial scale (regional scale vs. individual slope scale) and rock mechanical settings with respect to slope aspect, paraglacial bedrock instability may evolve rapidly or more linearly from glacial to non-glacial conditions. For explanation, see text.

10.5.2.3 Topo-climatic forcing on permafrost rockwalls

For the last two decades, a significant correlation of warm periods and intensified episodes of rock slope failures is found in the European Alps (Ravanel and Deline, 2015) and in the New Zealand Alps (Allen and Huggel, 2013). This trend is often associated with warming and thawing of bedrock permafrost, due to its high sensitivity to past and ongoing atmospheric warming (Gruber and Haeberli, 2007; Huggel *et al.*, 2012). During the 20th century, the mean annual temperature of mid-latitude Alpine settings has increased by more than 1 °C, especially since the Little Ice Age. For 2055, a warming of 2-3 °C (B1 global environmental scenario) or 3-4 °C (A1FI global economic scenario) is projected in comparison with the recorded 20th century average (Nogués-Bravo *et al.*, 2007). In the Turtmann Valley, the regional permafrost distribution was detected to be the most important factor controlling the spatial rockfall activity, indicated by the RF model (Fig. 10.6) as well as by the high ODs of PC1 and PC3 in the PCLR (Table 10.5). More than two thirds of rockfall source areas lie within a uniform altitudinal belt at ca. 2900-3300 m a.s.l., immediately above the predicted lower permafrost boundary (Fig. 10.5). In this zone, frozen bedrock plays a decisive role for slope stability, because changes of its thermal and hydrostatic state can significantly promote rockfalls (Krautblatter *et al.*, 2013). Seasonal and multi-annual freezing and thawing can lead to intense bedrock fracturing (Matsuoka and Murton, 2008), warming of subsurface temperatures may reduce the shear strength of ice-bonded discontinuities (Davies *et al.*, 2001) and water seepage from melting permafrost bodies can increase hydrostatic pressures (Gruber and Haeberli, 2007; Krautblatter and Hauck, 2007). Our findings are consistent with studies from the European Alps (Sass, 2005b; Ravanel and Deline, 2011; Fischer *et al.*, 2012), observing a similar altitudinal clustering of rockfall scars in a zonal window, where freeze-thaw cycles are highly effective. Assuming the existence of permafrost in our study area, it might be reasonable to infer that seasonal ice segregation in near-surface permafrost is one of the primary rock breakdown mechanisms in the hanging valleys. The bidirectional freezing progress over a period of several months and penetrates to depth of several meters (Hallet *et al.*, 1991; Murton *et al.*, 2006). This slow formation of segregation ice results in progressive fracture of the upper permafrost layer and the base of the active layer promoting rockfalls of low frequency, but of block sizes of decimeters to several meters in diameter. Even though data on frost cycles are missing in our study area, our assumption seems to be supported by field observations of the talus material along the permafrost-affected rock slopes (Fig. 10.8A). The obvious dominance of large block sizes might reflect the activity of large magnitude toppling events, which needs to be prepared by annual or decadal frost cycles with meter-scale freezing (Matsuoka and Murton, 2008). In contrast, rock fragments resulting from high-frequency diurnal frost cycles affecting the outermost

decimeters of bedrock tend to be much smaller and are found primarily near the apex of the talus slopes. Following Hales and Roering (2007, 2009), the regional-scale cracking intensity of segregation ice is likely highest in a rock temperature window between -3 and -8 °C, provided that water is available in the system, e.g. from the surface or from the active layer of permafrost. However, recent field studies of frost cracking inferred from acoustic emissions in natural Alpine rock-walls reveal slightly different temperature ranges. For instance, Amitrano *et al.* (2012) reported a rising of rock damage between 0 and -5 °C (granitic gneisses), while Girard *et al.* (2013) found frost cracking events occurring over the full range of temperatures from 0 down to -15 °C (crystalline rock). These field findings reveal that transferring the theoretical and laboratory-based frost cracking window concept (Walder and Hallet, 1985) to natural conditions is not trivial due to the mechanical and hydrological heterogeneity of bedrock. This is particularly the case when trying to upscale the frost cracking to larger spatial scales, where the relative importance of bedrock characteristics is difficult to evaluate or is simply underestimated. However, based on the datasets presented in this study, we cannot quantify the role of seasonal ice segregation opposite to other weathering processes such as daily frost action of the bedrock near-surface or thermal fatigue. To test our initial hypothesis that ice segregation is a major key driver for rockfall events, field data of bedrock temperature are needed in upcoming research to evaluate the penetration depth and timing of freeze-thaw cycles at the regional scale. Furthermore, the PCs of the LR model provide insights into the strength and direction of the interaction between bedrock permafrost and other factors governing rockfall activity. For instance, the factor loadings in PC3 (Table 10.5) support the notion that permafrost degradation through lateral heat fluxes is efficiently accelerated along convex topography, which is well exposed to high incoming solar radiation (Ravanel and Deline, 2011). However, the relative contribution of micro-scale bedrock roughness and curvature for the large-scale rockfall pattern in the Turtmann Valley was classified to be small (Fig. 10.6), contrary to the important role of overall valley topography. Permafrost-related rockfalls seem to be promoted most by steep terrain (up to 2.94 times, Table 10.5), linked to the higher shear stresses (Wyllie and Mah, 2004) and to the shorter penetration distance of the warming signal into the interior of the rock mass (Noetzli and Gruber, 2009). The critical slope threshold for failure of permafrost-affected bedrock was found to be $> 40^\circ$ (Fig. 10.5B), typical for Alpine rockfall events (Noetzli *et al.*, 2003). Furthermore, our results expose an aspect-driven rockfall activity. South-facing slopes show a decrease in rockfall activity. These slopes are characterised by higher solar radiation input and increased surface temperatures that increase the permafrost limit significantly compared to north-facing slopes (Table 10.3, PC3). Between N- and S-orientations, significant differences in mean annual rock surface temperature of up to 6 °C and in moisture supply are likely, as found by Coutard and Francou (1989) and Sass (2005a). The consequences for

frost weathering intensity (Hales and Roering, 2007) and resulting rockfall size and shape (Matsuoka, 2001; Hall and Thorn, 2011) are observable in the hanging valleys (Fig. 10.8). Although temperature fluctuations around 0 °C are probably highest at S-slopes, these dry thermal cycles might be only effective to weaken the outermost decimeter of bedrock, leading to the flaking off of small rock fragments (Hall and André, 2001; Prick, 2003). In contrast, seasonal freezing might advance much more slowly, but deeper along north-facing rockwalls up to several meters, significantly enhanced by the bidirectional moisture supply (Sass, 2005b; Murton *et al.*, 2006; Matsuoka and Murton, 2008). Consequently, rockfalls predominantly are expected to occur as large-magnitude, seasonal toppling events (Fig. 10.8A). Our study gives therefore evidence that the dependence of rockfall activity on aspect is probably not exclusively climatic, but rock mechanically predefined. Along north-facing slopes, in-dipping joints may favor moisture retention within the bedrock and allow subsurface advective heat fluxes to penetrate much faster and deeper through running melt water (Hasler *et al.*, 2011a,b). Besides predefining the block size and shape, the joint orientation with respect to aspect might significantly control the development of thaw corridors in permafrost and the efficiency of freeze-thaw cycles (Gruber and Haeberli, 2007). Therefore, our modeling results reveal that the influence of permafrost on rock slope stability can never be studied decoupled from other interacting factors, as its thermal and mechanical behaviour is considerably governed by the overall valley topography as well as small-scale morphometric and rock mechanical properties.

10.6 Perspectives

Coming back to Harrison (2001) stating that the relative role of rockfall controls can change with increasing scale due to emergent and complex system behaviour, leads us to the questions: "Would the findings of this study have been significantly different at the local scale?" And moreover, "would the key rockfall controls identified in our study area contrast with other mountain environments"? To examine these questions further research is needed in future. As each level requires its own scientific explanation, we cannot simply upscale local-scale knowledge and, in turn, we cannot down-scale the regional-scale knowledge obtained from this study to smaller phenomena. Thus, we hereby appeal to enlarge the research activities at each geomorphic scale. There is a need for both deductive studies to improve the mechanistic understanding of slope stability with respect to climate change and research at larger geomorphic scales; even if the latter may be associated to more narrative and abductive reasoning. A final goal must be an upscaling causation linking the local-scale rockfall controls to regional-scale conditions. Considering that the regional-scale importance of paraglacial and rock mechanical factors has been probably underestimated far too long, there is a par-

ticular need to consider more intensely those parameters in future geomorphic research. Here, we specifically aim to stress the great potential of the random forest algorithm for novel applications in studying complex, collinear geomorphic system behavior. Ideally, the classification results need to be tested against a method, which is not so much a black-box like the RF and which gives additional information on the strength of association between the interacting factors, such as a classical logistic regression using principal components. To overcome some sources of uncertainties, forthcoming work will comprise the validation of the lower permafrost boundary adjusted to the steep terrain and bedrock anisotropy by local field surveys. Similarly, local data of bedrock temperature are needed for more insight into the regional-scale intensity of frost cracking and its effective depth for rock breakdown. As a future step, it is desirable to test both our methods and the corresponding predictor variables in other mountain valleys where permafrost degradation and glacier retreat is dominant. Particularly in environments with vulnerable settlements and tourist infrastructure, using the approach developed in our study allows a relatively simple and time-efficient prediction of rockwalls most vulnerable to rockfall initiation. This in turn can be used as a key basis of information for regional hazard mitigation and sediment management in cold mountain regions that face ongoing atmospheric warming.

10.7 Conclusion

Rockfalls are among the most hazardous natural hazards and represent a first-order geomorphic agent in the sediment cascade of cold-mountain geosystems. Detailed knowledge on the critical factors controlling rockfalls is important for geomorphologists, engineers and decision makers. While most effort is obtained locally, the relative importance of potential rockfall controls at regional scales is still poorly understood. To bring further insight into the debate whether regional-scale rockfall activity is driven by (a) topo-climatic, (b) paraglacial (c) or rock mechanical factors, we designed a new spatial modelling approach using an inventory of 220 talus slopes in the Turtmann Valley (Swiss Alps). In this study, the classification-tree based random forest algorithm by Breiman (2001a) was applied for the first time for a rockfall-related purpose and combined with a classical logistic regression model using principal components. Major findings are:

- (a) The regional permafrost distribution was identified as the major control on the spatial rockfall activity. The clustering of rockfall source areas within a low-radiation altitudinal belt at 2900-3300 m a.s.l., consistent to the modelled permafrost probability of $> 90\%$, suggests that seasonal ice segregation growth in the near-surface permafrost might be one of the key rock breakdown mechanism. However, the relative contribution of permafrost strongly

depends on a complex interaction with small-scale bedrock morphometry and overall valley topography. Permafrost-affected rockfalls were found to be linked to convex, steep terrain ($> 40^\circ$) and north-facing valley flanks, promoting surficial moisture supply and subsurface lateral heat fluxes.

- (b) The paraglacial adjustment of rock slopes to the LGM glaciation and subsequent deglaciation was modelled to be the second most critical variable. Using ergodic reasoning, we detected an increasing rockfall probability with decreasing time since bedrock deglaciation. Besides glacially induced slope oversteepening relative to the specific rock mass strength, the gradual altitudinal rise of the frost cracking window during postglacial permafrost degradation is supposed to be the major large-scale paraglacial driver
- (c) The relative importance of rock mechanical properties for the regional-scale rockfall pattern was shown to be subdued compared to topo-climatic and paraglacial factors. Nevertheless, we proposed different hypothetical models, where the timing and intensity of paraglacial rockwall adjustment might evolve either exponentially or more linearly after LGM, dependent on the rock mass strength and the tectonically-derived dip direction of joints. Therefore, our study emphasises that periglacial rockfalls cannot be monocationally explained as they result from a complex synergetic interplay of topo-climatic, paraglacial and rock mechanical factors at different spatial and temporal scales. When evaluating bedrock instabilities with respect to past and recent deglaciation and permafrost degradation, our study demonstrates that it is never only the influence of changing climatic conditions, but rather the dependence on the topography and the structural geology of mountain valleys as well as the rock mass' memory effect on paraglacial forcing.

Acknowledgments We greatly appreciate the funding provided by the British Society for Geomorphology, the Humboldt-Ritter-Penck Foundation by the Gesellschaft für Erdkunde zu Berlin and the AK Geomorphologie. Furthermore, we thank Carina Schmitz for fruitful discussions on paraglacial geomorphology and her comments on the manuscript. We highly appreciate the previous achievements of the Research Training Group 437 – Landform, a structured and variable boundary layer at the Department of Geography, University (Otto and Dikau, 2010), especially Michael Nyenhuis for providing the permafrost modelling data. Finally, we would like to thank Sam McColl, Oliver Sass and the anonymous reviewer as well as the editors David Morche and, especially, Adrian Michael Harvey for their very constructive reviews that have significantly contributed to improve the manuscript.

Chapter 11

Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation

Hanna Meyer (1), Christoph Reudenbach (1), Tomislav Hengl (2),
Marwan Katurji (3), Thomas Nauß (1)

(1) Faculty of Geography, Philipps-University Marburg, Deutschhausstr. 10, 35037 Marburg,
Germany

(2) ISRIC — World Soil Information, P.O. Box 363, 6700 AJ Wageningen, The Netherlands

(3) Center for Atmospheric Research, University of Canterbury, Private Bag 4800, Christchurch
8020, New Zealand

Submitted to

Environmental Modelling & Software

Submitted 10 July 2017

11 Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation

Hanna Meyer, Christoph Reudenbach, Tomislav Hengl, Marwan Katurji, Thomas Nauss

Abstract

Importance of target-oriented validation strategies for spatio-temporal prediction models is illustrated using two case studies: (1) modelling of air temperature (T_{air}) in Antarctica, and (2) modelling of volumetric water content (VW) for the R.J. Cook Agronomy Farm, USA. Performance of a random k -fold cross-validation (CV) was compared to three target-oriented strategies: Leave-Location-Out (LLO), Leave-Time-Out (LTO), and Leave-Location-and-Time-Out (LLTO) CV. Results indicate that considerable differences between random k -fold ($R^2 = 0.9$ for T_{air} and 0.92 for VW) and target-oriented CV (LLO $R^2 = 0.24$ for T_{air} and 0.49 for VW) exist, highlighting the need for target-oriented validation to avoid an overoptimistic view on models. Differences between random k -fold and target-oriented CV indicate spatial over-fitting caused by misleading variables. To decrease over-fitting, a forward feature selection in conjunction with target-oriented CV is proposed. It decreased over-fitting and simultaneously improved target-oriented performances (LLO CV $R^2 = 0.47$ for T_{air} and 0.55 for VW).

keywords Cross-validation; Feature selection; Over-fitting, Random Forest; Spatio-temporal; Target-oriented validation

11.1 Introduction

Machine learning algorithms are well established in environmental sciences (Lary *et al.*, 2016; Kanevski *et al.*, 2009) and find application in a variety of fields as for example mapping of land cover (Ludwig *et al.*, 2016; Gislason *et al.*, 2006), vegetation characteristics (Lehnert *et al.*, 2015b; Verrelst *et al.*, 2012) and soil properties (Gasch *et al.*, 2015; Ließ *et al.*, 2016) as well as in geomorphological (Messenzehl *et al.*, 2017; Micheletti *et al.*, 2014) or climatological (Kühnlein *et al.*, 2014b; Hong *et al.*, 2004; Meyer *et al.*, 2016a; Appelhans *et al.*, 2015) studies. Most of the applications focus on static spatial predictions and are not aiming at estimating a certain variable simultaneously in space and time. However, though machine learning algorithms are still rarely applied in spatio-temporal models, the number of applications is increasing (Gokaraju *et al.*, 2011; Gasch *et al.*, 2015; Appelhans *et al.*, 2015; Meyer *et al.*, 2016b; Ho *et al.*, 2014; Jing *et al.*, 2016; Ke *et al.*, 2016; Lary *et al.*, 2014).

Machine learning algorithms in space-time applications learn from spatio-temporal observations to predict a certain variable for unknown locations and for an unknown point in time (within a defined model domain) allowing a monitoring of the environmental variable. The term “*prediction*”, in this context, should not to be confused with “*forecasting*” as most of the models are not aiming at predicting into the future but rather focus on predicting in past or present times as well as in space. In contrast to model-based geostatistics (Diggle and Ribeiro, 2007) as for example (co-)kriging, where one needs sufficiently distributed information on the variable at question for each interpolation time-step, spatio-temporal prediction models link a set of independent variables to the response (i.e. the variable in question) and only use those independent variables for the subsequent spatio-temporal prediction application. A typical example of spatio-temporal prediction models in environmental science might be the estimation of soil properties as done by Gasch *et al.* (2015). In this example, soil properties (volumetric water content, soil temperature and bulk electrical conductivity) are predicted in space and time on the basis of a machine learning model which is developed from a variety of spatial, temporal and spatio-temporal predictor variables as well as “*ground truth*” observations taken from data loggers.

Studies by Gasch *et al.* (2015) and Meyer *et al.* (2016a) have shown that the estimated performance of such models highly depends on the validation strategy: in both cases high differences between the performance estimated by a random test subset of the total dataset and the performance estimated by a Leave-Location-Out (LLO) Cross-Validation (CV) have been reported. LLO CV means that models are repeatedly trained by leaving the data from one location or a group of locations (i.e. climate stations, data loggers) out and using the respective held

back data for model validation. The differences between a random subset validation (lower error estimates) and LLO CV (higher error estimates) strongly suggest spatial over-fitting as the models can very well predict on subsets of the time series of the locations used for training, but fail in the prediction of unknown locations. The prediction on unknown locations, however, is in most cases the major task of such models. The LLO CV error must therefore be considered as the decisive performance indicator of spatial as well as spatio-temporal models. Similarly, spatio-temporal models have a risk of temporal over-fitting which needs to be assessed by Leave-Time-Out (LTO) CV (Gudmundsson and Seneviratne, 2015). However, it is these “*target-oriented*” validation strategies that focus on the model performance in the context of unknown space or unknown time steps that are not yet fully prevailed in literature. This is especially a problem as case studies ignoring the spatio-temporal dependence in the data have to be considered too optimistic (Roberts *et al.*, 2017). Even though LLO and LTO CV are used in some studies on spatial and spatio-temporal models (Ho *et al.*, 2014; Gudmundsson and Seneviratne, 2015; Ruß and Brenning, 2010; Meyer *et al.*, 2017c; Brenning *et al.*, 2012; Micheletti *et al.*, 2014), random k -fold CV, where the dataset is randomly partitioned into folds, is still considered common practice (Ke *et al.*, 2016; Messenzehl *et al.*, 2017; Ließ *et al.*, 2016; Ludwig *et al.*, 2016).

How to address spatial or spatio-temporal over-fitting in view to improved model selections? Over-fitting in machine learning models (when applied to spatial data) most likely happens due to poor representation of spatio-temporal sampling in predictor variable spaces. Hence, carefully selecting and interpreting predictor variables is a logical remedy for improving performance of spatial models. Many spatio-temporal prediction studies use auxiliary predictor variables which describe the properties of the location (e.g. elevation, slope, soil type, spatial coordinates). These variables vary in space but not in time which means that each station has a unique combination of static variables. We hypothesize hence that:

1. These temporally static variables are prone to over-fitting. Combinations of unique properties for each location are quasi comparable to a unique ID of the locations which is then used as predictor. Using such variables, the model is able to fit general characteristics of the individual time series.
2. Variables that lead to over-fitting can be automatically identified and removed using a feature selection method that accounts for the target-oriented performance.
3. Excluding misleading variables from the models does not only decrease over-fitting but also leads to improved target-oriented model performances.

Feature selection is an intuitive solution to reduce the number of variables to

the most important ones. However, the commonly used method for feature selection, Recursive Feature Elimination (RFE) (see e.g. Brungard *et al.*, 2015; Meyer *et al.*, 2017b,c; Ghosh and Joshi, 2014; Stevens *et al.*, 2013, in the field of environmental mapping), relies on variable importance scores which are calculated using solely the training subset (Kuhn and Johnson, 2013a). If a variable leads to considerable over-fitting, it has a high importance in the models. Therefore, this variable will be selected as important variable in the RFE process and is not removed regardless of a resulting high LLO CV error. Alternative approaches for detecting the over-fitting variables are hence required.

We consider two published case studies to demonstrate the effect of different validation strategies, the risk of spatial or spatio-temporal over-fitting as well as the potential of feature selection algorithms to minimize the degree of over-fitting. To estimate the degree of over-fitting, we compare the results of a random k -fold CV with the results of the target-oriented validation strategies LLO, LTO and Leave-Location-and-Time-out (LLTO) CV. We then compare the RFE method with a newly proposed forward feature selection (FFS) method that works in conjunction with target-oriented performance to identify and remove variables that lead to over-fitting. We implement all steps of data analysis and modeling in the R environment for statistical programming (R Core Team, 2016). Most of the analysis is based on the `caret` package (Kuhn, 2016b) that implements a wrapper to the Random Forest algorithm being used and provides functionality for data splitting and CV. All newly produced R functions and modeling steps are fully documented in <https://github.com/environmentalinformatics-marburg/CAST>

11.2 Case studies and description of the datasets

11.2.1 Case Study I: modelling air temperature in Antarctica

The first case study follows the approach of Meyer *et al.* (2016a) to spatio-temporally predict T_{air} in Antarctica based on LST data from the Moderate Resolution Imaging Spectroradiometer (MODIS) and auxiliary predictor variables. The dataset as it was used in the present study consists of 30666 hourly air temperature measurements from 32 weather stations distributed over Antarctica for the year 2013. The T_{air} values range from -78.40°C to 5.76°C with an average of -27.64°C and a standard deviation of 17.26°C .

Beside of MODIS based LST as a spatio-temporal predictor variable, several auxiliary spatial predictor variables were used that basically describe the terrain. In addition, a number of predictor variables that remain spatially constant but vary in time were used as temporal predictor variables. See Tab. 11.1 for the full list of predictors used in this study and Meyer *et al.* (2016a) for further

information on the dataset.

11.2.2 Case Study II: modelling volumetric water content of the "Cook-farm", USA

The second case study bases on the dataset applied in Gasch *et al.* (2015) to predict soil properties in 3D+time and can be freely accessed from the GSIF package in R. The research site of this case study is the R.J. Cook Agronomy Farm which is a 37 ha sized long-term agroecosystem research site in the Palouse region in the USA and operated by the Washington State University. The final dataset as prepared for this study consists of daily VW measurements from the years 2011 to 2013 taken by 5TE sensors (Decagon Devices, Inc., Pullman, Washington) initially installed in five depth (0.3, 0.6, 0.9, 1.2, and 1.5 m) at 42 locations within the study site. In this study we only focus on two dimensions plus time and limited the dataset to the depth of 0.3 m. The dataset then contained 33397 training samples. VW ranged from $0.093 \text{ m}^3/\text{m}^3$ to $0.613 \text{ m}^3/\text{m}^3$ with an average of $0.265 \text{ m}^3/\text{m}^3$ and a standard deviation of $0.076 \text{ m}^3/\text{m}^3$.

The covariables available from the research dataset that were used in this study as potential predictors to predict VW are a number of spatially continuous variables describing the terrain. Further, temporal variables as for example climate properties measured from the nearest meteorological station were used. See Tab. 11.1 for the full list of predictors used in this study and Gasch *et al.* (2015) for further information on the dataset.

11.3 Methods

11.3.1 Random Forest algorithm

Random Forest (Breiman, 2001a) was chosen as machine learning algorithm because it is a widely used algorithm, able to deal with both, numeric and categorical predictor variables, and because of its robustness to hyperparameter tuning (Kuhn and Johnson, 2013a). Random Forest bases on the concept of regression and classification trees, i.e. a series of nested decision rules for the predictors that determine the response. It repeatedly builds trees from random samples of the training data with each tree is a separate model of the ensemble. The estimations of all trees are finally averaged to produce the final estimate (Breiman, 2001a). To overcome correlation between trees, only a subset of predictors (`mtry`) is randomly selected at each split. The best predictor from the random subset is used at the respective split to partition the data. `mtry` is considered as a hyperparameter that needs to be tuned for a respective dataset in order to obtain an optimal

Table 11.1: *Predictor variables used within the two case studies with their dimension and resolution (res.). LST — Land Surface Temperature as measured by MODIS, Sensor - either MODIS Terra or Aqua, Ice — Ice covered ground or not, DEM - Digital elevation model, TWI — SAGA wetness index, NDRE.M — Normalized Difference Red Edge Index (mean), NDRE.sd - Normalized Difference Red Edge Index (s.d.), Bt — Occurrence of Bt horizon, BLD — Bulk density of soil, PHI — Soil pH, Precip_cum — Cumulative precipitation in mm, MaxT_wrcc - Maximum measured temperature, MinT_wrcc — Minimum measured temperature, Crop — Crop type. See also Meyer et al. (2016a) and Gasch et al. (2015) for further description.*

Case Study	Predictor	Dimension	Spatial res.	Temporal res.
<i>T_{air}</i> Antarctica	LST	2D+t	1km	instantaneous
	DEM	2D	1 km	-
	Aspect	2D	1 km	-
	Slope	2D	1 km	-
	Skyview	2D	1 km	-
	Ice	2D	1 km	-
	Sensor	(2D+t)	(1 km)	(instantaneous)
	Season	t	-	3 months
VW Cookfarm	Time	t	-	hour
	DEM	2D	10 m	-
	TWI	2D	10 m	-
	NDRE.M	2D	10 m	-
	NDRE.Sd	2D	10 m	-
	Bt	2D	10 m	-
	BLD	2D	10 m	-
	PHI	2D	10 m	-
	Precip_cum	t	-	1 day
	MaxT_wrcc	t	-	1 day
	MinT_wrcc	t	-	1 day
	Cdayt	t	-	1 day
	Crop	2D+t	10 m	1 year

trade-off between under- and over-fitting of the data. For a further description of Random Forest, see Breiman (2001a); James *et al.* (2013) and Kuhn and Johnson (2013a).

In this study, the Random Forest implementation of the `randomForest` package (Liaw and Wiener, 2002) in R was applied and accessed via the `caret` package (Kuhn, 2016b). Throughout the study, each Random Forest model consisted of 500 trees after no increase of performance could be observed using a higher number of trees. `mtry` was tuned for each value between two and the respective

number of predictor variables.

11.3.2 Validation strategies

To test the model performance on random subsets of the total datasets, a commonly used random 10-fold CV was used. Therefore, the data was split into 10 equally sized folds. Data splitting was done by stratified random sampling that ensures that the distribution of the response variable in each fold equals the distribution of the entire dataset. Models were then repeatedly trained by using the data of all except one fold and testing the model performance using the held-back data. In order to quantify the performance of the models using “*target-oriented*” validation strategies, the performance in view to the following criteria was tested (Fig. 11.1).

1. predict on unknown locations, tested by Leave-Location-Out Cross-Validation (LLO CV)
2. predict on unknown points in time, tested by Leave-Time-Out Cross-Validation (LTO CV)
3. predict on unknown locations and unknown points in time, tested by Leave-Time-and-Location-Out Cross-Validation (LLTO CV)

Therefore, the dataset was split into folds again, but this time each fold left the data of complete locations (LLO) or time steps (LTO) or locations as well as time steps (LLTO) out. For both case studies, the location of the data loggers defined a location and the dataset was split into 10 folds with respect to these locations. For LTO, the day of the year was used as splitting criterion for T_{air} Antarctica. For VW Cookfarm, data from more than one year was available allowing that individual months of each year could be left out for validation (12 months x 3 years = 36 unique time steps). Again, the data was split into 10 folds by leaving complete time steps out.

For all target-oriented validation strategies, the procedure was comparable to the random k -fold validation (which gives a biased estimate of prediction performance): models were repeatedly trained by using the data of all except one fold and testing the model performance for the held-back data. Over-fitting of the model in space and time was then quantified by comparing the random 10-fold CV results with the target-oriented validation results.

11.3.3 Feature selection

With the aim to remove predictors that are counterproductive in view to the target-oriented performance, we tested a RFE algorithm as well as a FFS

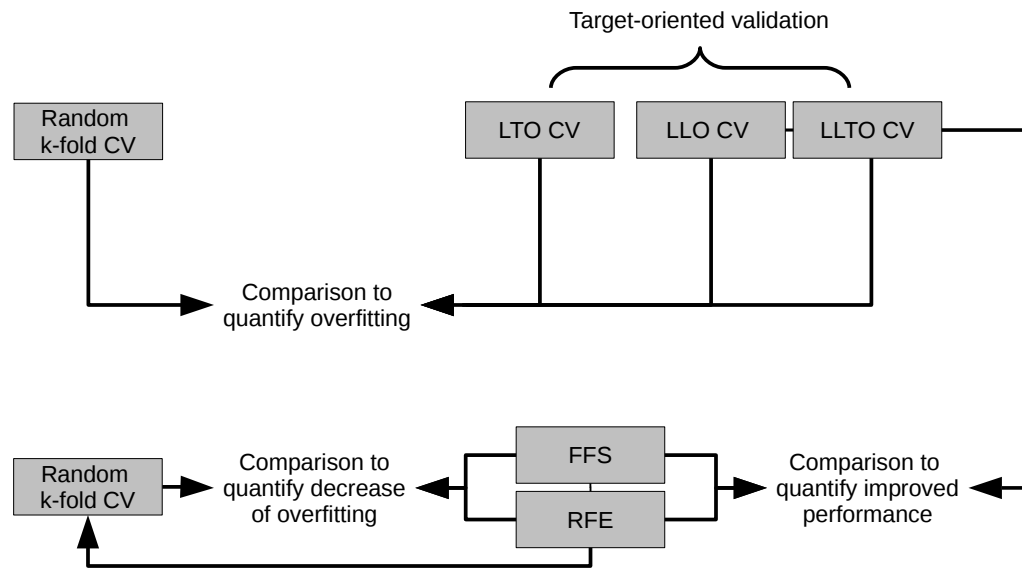


Figure 11.1: Schematic overview of validation strategies considered in this study. Leave-Location-Out (LLO), Leave-Time-Out (LTO) and Leave-Location-and-Time-Out (LLTO) cross-validations (CV) were used as target-oriented strategies. LLO and LLTO CV were used in conjunction with recursive feature elimination (RFE) and forward feature selection (FFS) to reduce spatial over-fitting and its impact.

algorithm that works in conjunction with target-oriented validation (Fig. 11.1). We used LLO and LLTO CV as target-oriented validation strategies as the ability of the model to predict on unknown locations was of upmost importance for both case studies.

RFE relies on variable importance scores that are calculated during the initial random forest model training. The algorithm successively removes the least important variables to find the best performing set (see Kuhn and Johnson, 2013a, for further details). In this study, we used the RFE implementation from the caret package (Kuhn, 2016b). As outlined above, we assume that RFE is not a helpful approach to overcome spatio-temporal over-fitting as variables are not ranked according to target-oriented performance. As an alternative approach, we developed and implemented a FFS algorithm in R (Algorithm 1). The algorithm first trains models (i.e. Random Forest) of all possible 2-variable combinations of the total set of predictor variables. The best initial model in view to target-oriented performance is kept. The number of predictor variables is then iteratively increased. The improvement of the model is tested for each additional predictor using target-oriented CV. The process stops when none of the remaining variables decreases the error of the currently best model. The algorithm therefore fits a maximum of $2 * (n - 1)^2 / 2$ models (e.g. 81 models when 10 predictors are considered).

```

for each resampling iteration do
  Partition the data into training and test data;
  Tune and train models using all possible 2-variable combinations;
  Predict on test data and calculate model performance;
end
Keep the best performing 2-variable model ( $model_{best}$ );
for each additional number of variables  $i$ ,  $i=3...N$  do
  for each remaining variable  $V_R$  do
    for each resampling iteration do
      Partition the data into training and test data;
      Tune and train models using the variables of  $model_{best}$  and  $V_R$ ;
      Predict on test data and calculate model performance
    end
  end
  if  $mean(error\ of\ model_i) > mean(error\ of\ model_{best})$  then
    break
  end
  Keep the best performing  $i$ -variable model ( $model_{best}$ );
end

```

Algorithm 1: Pseudo-code description (similar to the ones from the caret package) for the FFS algorithm. Resampling in this study bases either on LLO or LLTO

11.4 Results and Discussions

11.4.1 Target-oriented validation

For both case studies, a random k -fold CV showed a high performance with only low differences between observed and predicted values, indicating a nearly "perfect fit" of the data (model $T_{air}/VW01$ in Tab. 11.2). However, in view to unknown locations (LLO CV), the performance decreased considerably (models $T_{air}/VW02$ compared to models $T_{air}/VW01$ in Tab. 11.2). This means that the model was generally less able to predict beyond the location of the training data compared to what might have been expected regarding the random k -fold CV error. The ability of the T_{air} model to predict the outcome for an unknown day within the temporal model domain of 2013 remained high (model $T_{air}03$ in Tab. 11.2). Thus, in view to unknown locations and unknown days (model $T_{air}04$ in Tab. 11.2), the error was comparable to the LLO CV error. Uncertainties in view to unknown locations were the major source of error. The temporal error had more effect on the VW Cookfarm example where complete months were left out for validation (model VW03 and VW04 in Tab. 11.2).

Since the differences between random k -fold CV and target-oriented CV are noticeably high, the results highlight the need to perform CV in view to the model target in order to draw meaningful conclusions. If the aim is to map the response variable, one must consider LLO CV as decisive error indicator as

the random CV error can lead to considerable misinterpretations of the model performance. Especially when the model is to be applied on unknown years, the potential of the model to predict beyond the years used for model training must also be considered. In this case, LLTO CV can assess the error in both, space and time, however, the number of validation data decreases as the overlap between LLO and LTO is used. This causes the results to be less robust compared to a separate view on LLO and LTO CV where more data are available for testing.

11.4.2 Detecting over-fitting

As the models performed well on random subsets of the entire datasets (random k -fold CV) but had high errors when faced with unknown locations, spatial over-fitting must be suspected for both case studies. The model could only lead to high performances when information about a respective location went into model training. Therefore, the model was over-fitting in space as only locations used for training could reliably be predicted by the model. Subsequently, also LLTO CV showed high errors, though temporal over-fitting only slightly contributed to that error in case of the T_{air} Antarctica example. In this case study, over-fitting in time was a minor issue, at least on the considered time scale (days). In the case study of VW Cookfarm, the time scale used for data splitting was months of the individual years. Considering these larger time scales that were left out, the model performance decreased compared to the random k -fold CV performance ($R^2 = 0.79$ compared to 0.92, see VW03,01 in Tab. 11.2). Thus, temporal over-fitting must be assumed in addition to spatial over-fitting as only months that went into model training could reliably be predicted by the model.

11.4.3 Reducing over-fitting and improving model performances

To decrease the impact of over-fitting, RFE and the newly designed FFS were compared. On the first sight, RFE reduced over-fitting in the T_{air} Antarctica example, getting obvious in lower differences between random k -fold CV and target-oriented CV (Fig. 11.2a, model $T_{air}05$ compared to $T_{air}06$ as well as $T_{air}09$ compared to $T_{air}10$ in Tab. 11.2). This pattern, however, could not be supported by the VW Cookfarm example, where the differences between random k -fold CV and target-oriented CV remained equally high (Fig. 11.2b, model VW05 compared to VW06 as well as VW09 compared to VW10 in Tab. 11.2). In fact, this was the expected pattern as the variable importance ranking within the RFE is based on internal importance estimates (Fig. 11.3) without consideration of the importance in view to target-oriented errors.

The explanation for the effect shown in the T_{air} example lies in the ranking of the variables (Fig. 11.3a): Among the most important variables were appar-

Table 11.2: Regression statistics between observed and predicted values of air temperature (T_{air}) and volumetric water content (VW) based on cross-validation (CV). Models were validated using random k-fold or using target-oriented Leave-Location-Out (LLO), Leave-Time-Out (LTO) and Leave-Location-and-Time-Out (LLTO) CV. Recursive feature elimination (RFE) and the newly proposed forward feature selection (FFS) were tested. Performance measures are mean error (ME), mean absolute error (MAE), root-mean-square-error (RMSE) and coefficient of determination (R^2). Bold numbers indicate the decisive objective error estimates after misleading variables were removed by FFS. Compare target-oriented CV without feature selection to random k-fold CV to estimate over-fitting. Compare LLO and LLTO CV using RFE or FFS to estimate the increase of performance compared to LLO and LLTO CV without feature selection. Note that the random CV performance is only provided for comparison but cannot be regarded as a meaningful measure.

Model	CV	Feature Select.	ME	MAE	RMSE	R^2
$T_{air}01$	random	none	0.016	4.155	5.556	0.899
$T_{air}02$	LLO	none	0.068	12.178	15.850	0.244
$T_{air}03$	LTO	none	0.017	4.244	5.665	0.894
$T_{air}04$	LLTO	none	0.236	12.164	15.807	0.246
$T_{air}05$	LLO	RFE	0.011	10.353	13.647	0.400
$T_{air}06$	random	variables of $T_{air}05$	0.025	9.113	12.021	0.519
$T_{air}07$	LLO	FFS	0.072	9.756	12.564	0.474
$T_{air}08$	random	variables of $T_{air}07$	0.000	8.602	11.157	0.583
$T_{air}09$	LLTO	RFE	0.405	10.251	13.416	0.413
$T_{air}10$	random	variables of $T_{air}09$	0.025	9.113	12.021	0.519
$T_{air}11$	LLTO	FFS	0.253	9.658	12.387	0.485
$T_{air}12$	random	variables of $T_{air}11$	-0.001	8.601	11.156	0.583
VW01	random	none	0.00	0.016	0.023	0.919
VW02	LLO	none	-0.002	0.041	0.054	0.488
VW03	LTO	none	-0.001	0.024	0.035	0.794
VW04	LLTO	none	-0.007	0.040	0.050	0.500
VW05	LLO	RFE	-0.002	0.041	0.055	0.475
VW06	random	variables of VW05	0.000	0.014	0.021	0.931
VW07	LLO	FFS	0.000	0.037	0.051	0.552
VW08	random	variables of VW07	0.000	0.036	0.049	0.580
VW09	LLTO	RFE	-0.007	0.040	0.050	0.502
VW10	random	variables of VW09	0.00	0.015	0.022	0.926
VW11	LLTO	FFS	-0.004	0.039	0.051	0.499
VW12	random	variables of VW11	0.00	0.036	0.049	0.580

ently those that do not lead to an over-fitting. Only the top three variables were selected by the RFE ("season", "time" and "LST", see Tab. 11.1 for explanation) that could in this example lead to a reduced effect of over-fitting. Including just one additional variable (in this case "aspect" as this was the variable rated as next important, see Fig. 11.3) was recognised as counterproductive by the RFE. However, the example of VW Cookfarm demonstrates that this pattern is rather chance than a systematic ability of the RFE design to remove over-fitting variables. In the VW Cookfarm example the variables that were ranked as important led to over-fitting so that the RFE could not decrease this problem by removing least important variables. Over-fitting in this example is generated because the variables were not ranked according to their target-oriented importance within the models. In fact, the RFE algorithm kept all except three variables thus it yielded the best performance using nearly the full set of predictors which, however, could not remove over-fitting.

The FFS algorithm, in contrast, could reliably reduce the differences between random k -fold CV and LLO as well as LLTO CV in both case studies: when the respective less-variable model was validated with random k -fold CV, the differences to the LLO as well as LLTO CV error decreased (Fig. 11.2, model VW/ T_{air} 07,11 compared to VW/ T_{air} 10,12 in Tab. 11.2). This shows that removing misleading variables decreased the problem of spatial over-fitting. In the case study of T_{air} Antarctica, it suggested the combination of "season", "ice", "LST", "sensor", "aspect" as necessary variables and rated all others as counterproductive. For VW Cookfarm, the variables "Precip_cum", "cdayt", "MaxT_wrcc", "MinT_wrcc", "Crop" were suggested to yield optimal results in view to LLO as well as LLTO CV.

The variables that were rated as counterproductive and have been removed during FFS were mainly spatially continuous but temporally constant variables. Especially in the case study of T_{air} Antarctica, such variables formed a distinct "pointer" on the individual logger locations, as each logger location featured unique combinations of the spatial variables (i.e. unique combinations of slope, aspect, altitude). Therefore, these variables are, in combination, comparable to an "ID" for the loggers that was then used as predictor. ID-like predictors enable the algorithms to access individual characteristics of the time series of the loggers which in turn leads to a misinterpretation of such variables: these variables are associated with logger-specific patterns that cover the true underlying relations between these predictors and the response. This suspicion is supported by a high internal importance of such variables within the models (Fig. 11.3) especially in the VW Cookfarm example (e.g. NDRE.M+BLD+PHI) but a removal of these variables during the FFS. Under these considerations, the behaviour of the RFE to reduce the impact of over-fitting in the T_{air} Antarctica example becomes understandable: as the top ranked variables "season", "time" and "LST" are not

prone to spatial over-fitting, the RFE could yield best results using only these three variables. If an over-fitting variable was amongst the top two variables, over-fitting could not have been resolved and the differences between random k -fold and target-oriented CV errors stayed high as in the VW Cookfarm example.

Removing counterproductive variables using FFS did not only lead to reduced over-fitting but also to improved target-oriented performances (Fig. 11.2, 11.4, Tab. 11.2). This is especially obvious for the T_{air} Antarctica data where the LLO CV R^2 increased from 0.24 to 0.47 (model $T_{air}07$ compared to $T_{air}02$ in Tab. 11.2, Fig. 11.2a, Fig. 11.4a). The patterns for LLTO CV were the same (model $T_{air}11$ compared to $T_{air}04$ in Tab. 11.2, Fig. 11.4a). Also in the VW Cookfarm example FFS led to an increased LLO performance, though the effect was less strong compared to the T_{air} Antarctica data (Fig. 11.4b). The LLO CV R^2 increased from 0.49 to 0.55 (model VW02 compared to VW07 in Tab. 11.2) using only the selected variables. For the LLTO CV error, the FFS did not result in an improved model performance (model VW04 compared to VW11 in Tab. 11.2, Fig. 11.4b) though over-fitting could be significantly removed (Fig. 11.2b). Obviously removing misinterpreted variables could not improve the performance which suggests that the potential of the variables to predict beyond the training locations and months is depleted. However, this model is now more robust as only a small subset of the initial variables are used and over-fitting could be reduced.

Though FFS is time consuming, it is able to automatically detect and remove variables that are counterproductive in view to the target. The computation time can be decreased by thorough pre-selection of potential predictors in view to their effect in space and time to avoid ID-like pointers on individual locations or time steps. Considering the potential of FFS as shown in this study to remove counterproductive variables in view to a target-oriented performance, it is likely that the algorithm is able to improve a variety of published models beside of the two case studies (Meyer *et al.*, 2016a; Gasch *et al.*, 2015). As an example, Langella *et al.* (2010); Shi *et al.* (2015) and Janatian *et al.* (2017) used latitude and longitude as predictors which are prone to create an ID of the locations used for training.

The focus of this study was on spatio-temporal models, however, most of the findings apply for purely spatial models as well. This is supported by the studies of e.g. Micheletti *et al.* (2014) and Roberts *et al.* (2017) who left spatial units out for validation and yielded less optimistic results compared to a random k -fold CV, thus spatial over-fitting is indicated. Also Li *et al.* (2011) included latitude and longitude as predictors in a purely spatial model and observed linear features in the resulting map. If such models are validated with random k -fold CV, a statistically good fit is feigned but spatial over-fitting occurs as a consequence of the misinterpretation of certain variables. In such applications, the proposed FFS

in conjunction with target-oriented validation (in this case leave-spatial-unit-out CV) can improve the model results and will produce more robust results.

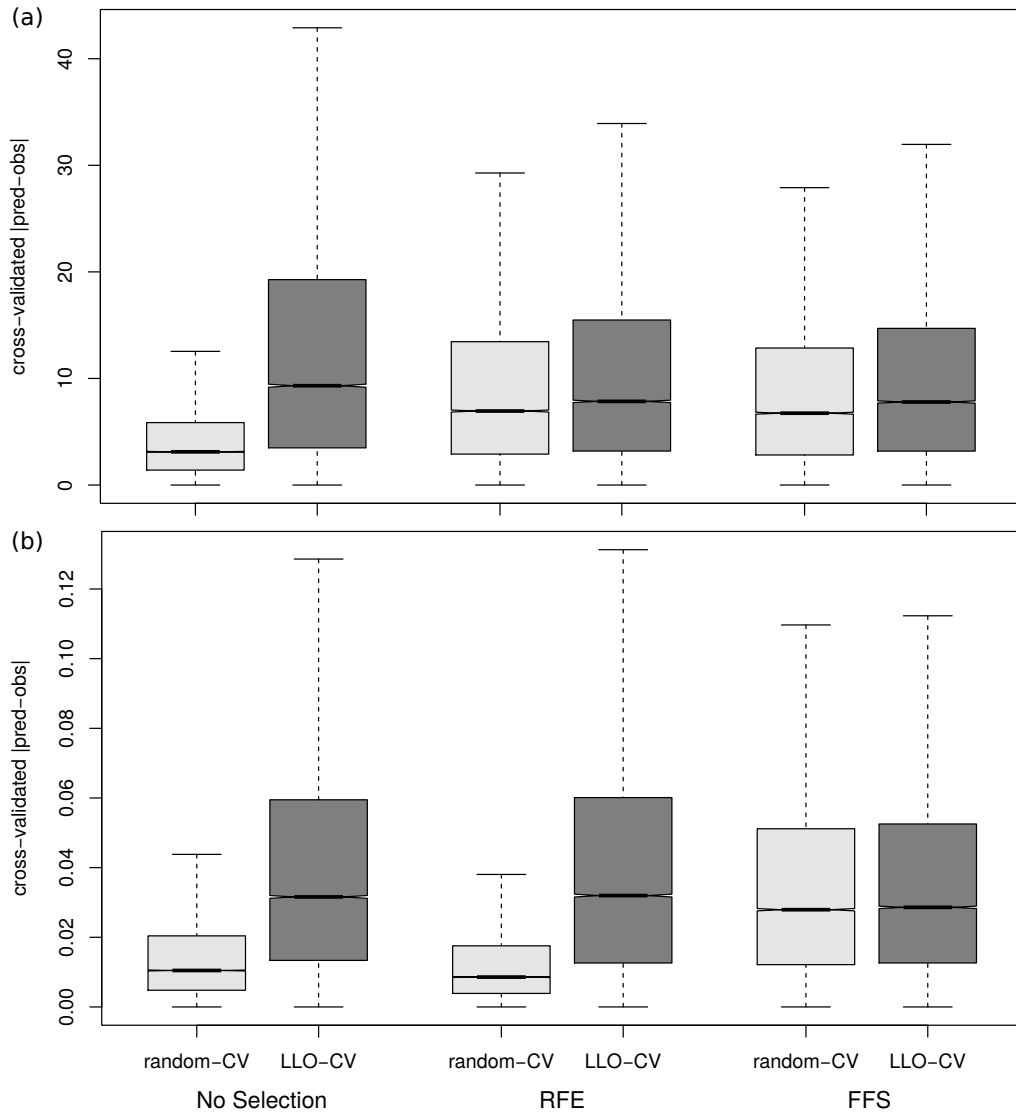


Figure 11.2: Differences in the Leave-Location-Out (LLO) cross-validation (CV) performance of a) the air temperature (T_{air}) estimations and b) the volumetric water content (VW) estimations using different feature selection strategies. The effect of a Recursive feature elimination (RFE) and the newly proposed forward feature selection (FFS) are compared. The variables selected by RFE for the case study of T_{air} Antarctica were "season", "time", "LST". FFS selected "month", "ice", "LST", "season", "sensor". For the case study of VW Cookfarm all potential predictors except "Bt", "TWI", "MinT_wrcc" were selected by the RFE. FFS selected "MaxT_wrcc", "cdlayt", "Precip_cum", "Crop", "MinT_wrcc". See Tab. 11.1 for further explanations on the variables.

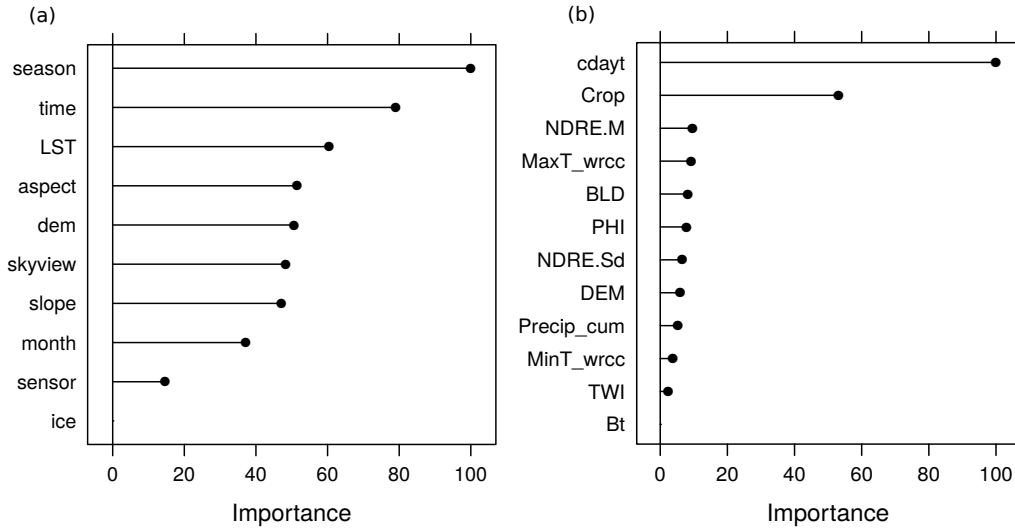


Figure 11.3: Relative scaled importance of the predictor variables within the Random Forest models for the case study of (a) T_{air} Antarctica and (b) VW Cookfarm. See Tab. 11.1 for further explanations on the variables

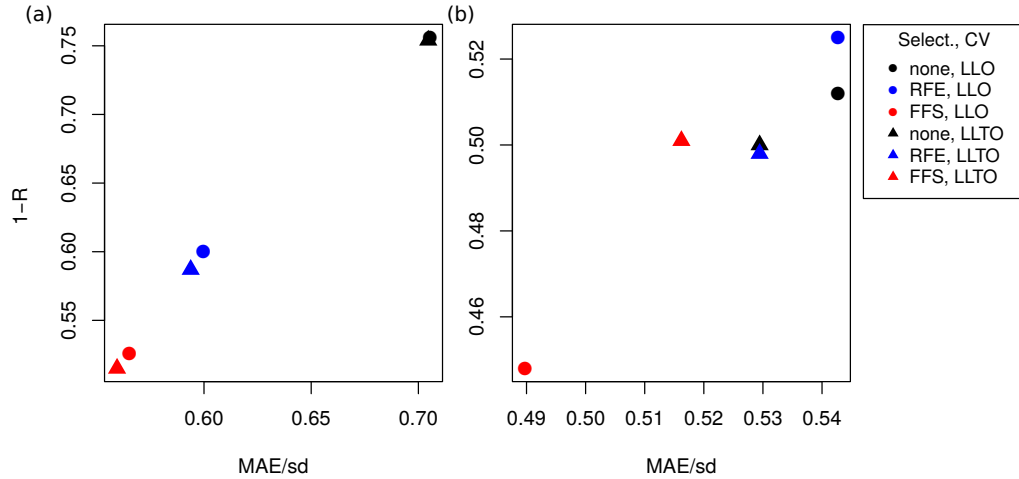


Figure 11.4: Differences in the Leave-Location-Out (LLO) and Leave-Location-and-Time-Out (LLTO) cross-validation (CV) performance using no feature selection, a recursive feature elimination (RFE) and the newly proposed forward feature selection (FFS) of the a) air temperature (T_{air}) Antarctica models and b) the volumetric water content (VW) Cookfarm models. Performance is indicated using the mean absolute error (MAE) divided by the standard deviation (sd) of the mean, and the proportion of variation unexplained ($1 - R^2$). Colors indicate the different feature selection strategies. The shape indicates the CV method being used. Performance increases from the upper right corner towards the lower left corner. It is shown that models using no feature selection generally have the lowest performance and models using FFS have the highest performance.

11.5 Conclusions

This study aimed at demonstrating the effect of target-oriented validation and at finding a solution to detect and reduce spatial over-fitting. For this we used two previously published case studies. We discovered high differences between random k -fold and target-oriented CV: the random k -fold CV R^2 of the T_{air} Antarctica study was 0.90, contrasting to a LLO R^2 of 0.24 and the random k -fold CV R^2 of the VW Cookfarm study was 0.92 compared to a LLO CV R^2 of 0.49. This shows that errors estimated with a standard random k -fold CV can considerably deviate from target-oriented error estimates which highlights the clear need for target-oriented validation to avoid an overoptimistic view on results.

We further hypothesized that the observed patterns of spatio-temporal over-fitting are caused by temporally constant predictors (e.g. elevation, slope, ...) that act in conjunction with each other like an ID. This occurs when locations used for model training have unique spatial properties. It appears that the models of both case studies were able to learn general characteristics of the time series of the individual locations. The models were then very well able to predict subsets of the time series (low random k -fold CV error), but then failed to predict beyond the training locations (high LL(T)O CV error). To automatically detect and remove variables that lead to over-fitting, we proposed using the FFS algorithm in conjunction with target-oriented validation. By removing the misleading predictors, the FFS was able to automatically reduce spatio-temporal over-fitting which was reflected in similar errors for random k -fold CV and target-oriented CV. After refitting the models using the FFS procedure, the LLO CV R^2 was 0.47 for T_{air} Antarctica and 0.55 for VW Cookfarm, hence the proposed method could improve the target-oriented model performance.

Though predicting environmental variables in space and time remains challenging, validation strategies suggested in this article allow assessing model errors objectively and allow identifying over-fitting. Despite the general opinion that Random Forests are insensitive to over-fitting, unfavorable combinations of predictors and/or distribution of the training data in space and time can lead to serious over-fitting effects. In this study, variables that has caused that over-fitting were removed from the models and the model performance has immediately improved. However, certain variables might be misleading but still contain valuable information. How to minimize the over-fitting effect of such variables but still use them in the spatial prediction, remains to be solved. With an increasing application of machine learning for spatio-temporal predictions, further studies and procedures for preventing over-fitting in machine learning applications will hence be increasingly important.

Acknowledgments This work was partly funded by the Federal Ministry of Education and Research (BMBF) within the IDESSA project (grant no. 01LL1301) which is part of the SPACES-program (Science Partnership for the Assessment of Complex Earth System processes). It was further supported by the Ross Sea Region Terrestrial Data Analysis research program, funded by the Ministry of Business and Innovation, New Zealand, with contract number CO9X1413.

Chapter 12

Conclusions

12 Conclusions

The individual chapters contain various contributions to an enhanced understanding of machine learning for environmental geography, and for advancing the respective scientific field. Thus, each chapter includes particular methodological challenges and/or new developments as well as new scientific findings which was possible only by the combined development of methodological and scientific knowledge. The in-depth study of machine learning applications in different thematic fields as outlined in this thesis, allowed drawing conclusions and delineate important consequences for its applicability in environmental geography. In this concluding chapter, the main methodological and scientific outcomes in general will be emphasized with special focus on the challenges and opportunities of machine learning in view to spatio-temporal data.

12.1 Significance of the developed products

Within this thesis, several scientifically relevant products have been developed based on the utilization of machine learning. Special focus in the thesis was on the monitoring of rainfall dynamics (chapters 2, 3, 4). Since rainfall is a difficult parameter to assess, its retrieval required methodological pre-studies before a model could successfully generate a monitoring product for southern Africa that was representative for areas with low densities of climate stations. In this context, a set of most adequate machine learning algorithms was tested to identify the best performing algorithm (chapter 2). However, since no significant differences between the tested algorithms could be found, the first important finding was that an optimization of the rainfall retrieval must rely on the choice of optimal predictor variables rather than on the choice of the machine learning algorithm. Consequently, more in-depth scientific considerations on potential predictors for rainfall were the objectives of the subsequent study (chapter 3). The findings from both studies provide an improved methodological framework that allowed estimating rainfall for areas with a low density of climate stations, such as southern Africa (chapter 4). Though monitoring of rainfall remains highly challenging and the newly derived spatio-temporal rainfall product is still afflicted with uncertainties, the product outperforms a sophisticated global rainfall product which emphasizes the great potential of the methods applied in this study. The newly developed monitoring product is of high relevance for farmers in southern Africa who vitally depend on rainfall. It is also most important for applied science since rainfall represents a key parameter for ecological and hydrological studies.

Also the new air temperature monitoring product for Antarctica, which was developed in this study (chapter 5), is an important baseline product for ongoing

research. With the high resolution spatio-temporal air temperatures of entire Antarctica it provides for the first time a reliable temperature product with an objective error estimation (see section 12.4) under challenging regional conditions. The result can be considered a breakthrough for highly heterogeneous areas, such as the Mc Murdo Dry Valleys, where sub-regional environmental processes are extensively studied but high resolution air temperature datasets were missing.

In view to biogeographical applications, this thesis considered different spatial and spectral scales to move beyond point measurements towards larger scale mapping or monitoring. Chapter 6 tested the significance of hyperspectral point measurements in order to get proxies of pasture degradation for the Tibetan Plateau. In this context, it could be shown that multispectral information was sufficient to estimate vegetation cover and biomass and hyperspectral data could not improve the estimations. This is an important finding as it facilitates work since more multispectral satellite data are available and its analysis far less complex compared to hyperspectral data. Spectral resolution as well as spatial scale was also relevant in chapter 7 where high spatial resolution Google Earth RGB images of South African Savannas were classified into woody and non woody vegetation. The classified images form a database of new ground truth data that serve as a baseline for a planned larger scale spatio-temporal monitoring of bush encroachment. Starting from very high resolution is important as the vegetation in the savannas appears patchy and lower resolution satellite pixels therefore contain mixed signals of different vegetation. Creating spatially explicit high resolution ground truth data allows accounting for local variability but transferability to larger scales.

So far, a variety of studies was presented where environmental characteristics were either modelled in 2D (chapters 6, 7) or 2D+time (chapters 2, 3, 4, 5). Chapter 8 extends the modelled dimensionality at the example of soil properties (moisture, temperature, electrical conductivity) by considering the depth as an additional dimension. This chapter represents the so far first study of a 3D+time model in environmental science. The results of this chapter are therefore not only important as a baseline to understand spatio-temporal processes in soils but are especially valuable from a methodological perspective as they form a framework for 3D+time modelling of the environment. Chapter 9 also deals with modelling of soil properties, specifically soil respiration. As soil respiration is an important factor in carbon dynamics, knowledge about its spatio-temporal patterns are important in view to assess the release of carbon from the soils e.g. under different land use management practices. This study is focussing the applied aspects of the newly developed methodological framework as derived from chapters 2-8. A model was to be developed that allows estimating soil respiration under different soil moisture and temperature conditions based on mid infrared spectroscopy. The developed model allows avoiding complicated and

time consuming laboratory work in order to assess soil respiration. Therefore, the model itself is the valuable outcome as it allows scientists to estimate soil respiration from MIRS data solely.

Machine learning, however, cannot only be used to create spatio-temporal estimates of the environment, but it further gives important information about the explanatory power of its drivers. Though machine learning is considered as a "black-box" and exact relationships between the predictors and the response cannot be retrieved, chapter 10 shows at the example of rockfall modelling that nevertheless machine learning still has great potential for contributing to the understanding of the driving forces of environmental processes.

12.2 Opportunities of machine learning in geography

With the ongoing popularity of data-driven approaches, it is sometimes criticized that data-driven data analysis is the "death" of knowledge-based science (Mazzocchi, 2015; Anderson, 2008). However, as a result of the in-depth study of the applicability of machine learning in different fields of environmental geography, this thesis emphasizes that data-driven science presents greater opportunity rather than a risk. Comprehensive data-driven science is not, as feared, accompanied by a loss of understanding of underlying environmental processes. In contrast, it could be shown that it is crucial to have a considerable understanding of ecosystems and processes combined with strong methodological knowledge to solve complex problems. In this context, data-driven science along with machine learning offers opportunities that simply cannot be achieved using traditional methods. Machine learning allows the inclusion of large numbers of predictors with the possibilities to include different types of variables as numeric, ordinal or nominal predictors. This is an advantage where common approaches in respective fields (e.g. partial least squares regression, chapter 9) or stepwise logistic regression, chapter 10) failed as incorporating non numeric information is challenging. However, the most important advantage of machine learning is not the option to include various types of variables but that no *a priori* assumption about the relationships between variables is required. This allows integrating a wide range of data with completely different relationships. This becomes especially important when large numbers of variables are used (e.g. chapters 3, 6) and individual relationships are difficult to assess due to the volume of data. Considering these characteristics, it is not surprising that machine learning algorithms have been shown to be superior compared to traditional methods (chapters 5, 8, 10). In view to the ongoing trend from small, towards big geographic data (Miller and Goodchild, 2015), machine learning must therefore be considered as highly relevant for geography offering great opportunities in view of modelling complex

systems.

12.3 Challenges of machine learning in geography

In the previous section it has been indicated that machine learning offers great flexibility and low restrictions for model development. Admittedly, this freedom comes with a risk that makes the mentioned critical attitude towards data-driven science understandable. Algorithms that can handle all kinds of variables and arbitrary relationships that are learned by the algorithm itself naturally go in hand with the risk of a "blind" model application without a critical confrontation with the environmental system or the method itself. This is especially problematic as current software developments allow machine learning to be applied by virtually everyone. Non comprehensively applied machine learning algorithms lead to considerable misinterpretation of the results (see chapters 5, 11) and don't advance, but rather harm, the scientific fields. Therefore, the opportunities of machine learning can only be exploited if the methods are applied with care.

There are several characteristics of geographic data (see chapter 1) that create challenges that need to be considered in view of a comprehensive application of machine learning. Some of them, for example handling unbalanced data, can be resolved through a thinking beyond the "standard application" and will be summarized in section 12.5. However, some challenges that were addressed in this thesis, are more challenging, requiring very thoughtful consideration and will certainly still be challenges for ongoing research.

Scale dependency makes machine learning applications challenging in geography, especially in remote sensing applications. It is optimal if the ground truth data can be sampled on the same spatial resolution as the predictors (e.g. by classifying high resolution images as a reference, see chapter 7). Remote sensing-based predictor variables usually don't share the same spatial scale as ground truth data that are sampled at small spatial extents in the field (e.g. chapters 4, 5, 6). This bears the risk that the ground truths reflect local characteristics but the predictors reflect the general signal of a larger area. Therefore, studies must act on the assumption that the reference data are representative for the entire pixel of the predictor variables. Field work or the selection of available ground truth data must consequently be in view to the spatial resolution of the predictor variables.

Certainly, the major challenge of machine learning in spatial or spatio-temporal applications, is the consideration of spatial and temporal dependencies in the data. This point is extensively addressed in chapter 11 and will be summarized in the following (section 12.4).

12.4 Accounting for space-time dependencies

The problem of incorporating space-time dependencies in machine learning applications was identified as the major challenge which bears high risks for considerable misinterpretations. Chapter 11 picked up on the issues associated with space-time dependency, firstly mentioned in chapter 5.

A key characteristic of geographic data is their spatial and/or temporal autocorrelation (chapter 1), thus, geographic data are dependent in space and time. Spatio-temporal autocorrelation becomes obvious considering the example of air temperature measurements by climate stations discussed in chapter 5: measured air temperature from a climate station is highly temporally autocorrelated as the temperature measured now is related to the temperature measured an hour ago. Similarly, the data are autocorrelated in space as a neighbouring station is likely to feature a similar air temperature. When air temperature is modelled using machine learning, random k-fold CV is the most commonly used model tuning and validation strategy. However, this validation strategy answers the question of how well the model performs on a random subset of the data. This makes only sense if the data are independent from each other, but it is not meaningful for geographic data with spatial and temporal dependencies. As soon as there is a dependency, a random subset validation will generate overoptimistic results as dependent data are used for model training as well as validation (e.g. same day or same location). Therefore, random k-fold CV can only be considered as a meaningful strategy if the samples are independent from each other and does not answer our questions. What we want to know in our example is how well the model performs for an unknown location or for an unknown point in time (both within a spatio-temporally defined model domain) in order to assess the model's ability to be applied in larger spatial or temporal contexts. Therefore, we need target-oriented validation strategies that address exactly these questions by applying LLO, LTO or even LLTO CV strategies rather than a random k-fold CV (chapter 11). Though this seems to be obvious, it is not yet common practice in geographic machine learning applications and need to be enforced in view to objective model validations and comparisons. The need of a target-oriented validation is particularly obvious for space-time data that are affected by both, spatial and temporal autocorrelation. However, it is also important for spatial mapping without a temporal component to account for the spatial dependencies (chapter 6).

Though the target-oriented validation strategies allow for objective error assessment, it became obvious (firstly mentioned in chapter 8) that the error estimates are often considerably lower when validated with random k-fold CV compared to target-oriented validation. This pattern could firstly be attributed to

spatial over-fitting in chapter 5. It could be shown that spatio-temporal dependencies caused certain variables to be misinterpreted by the algorithm. To identify and remove misleading information, standard methods (e.g. recursive feature selection) are not appropriate as they are not designed to work in conjunction with target-oriented validation. Therefore, a new feature selection was developed in this thesis (chapters 5, 11) that allowed for an identification of misleading variables. The optimized method could reduce over-fitting and led to improved statistical model performances. It therefore allows for a more realistic model which can be applied on larger spatio-temporal scales.

12.5 Practical consequences

As stated, one needs to think beyond the "standard way" of machine learning in order to take practical advantage for the spatial environmental sciences. From the experience gained in this thesis, the following section summarizes some general methodological consequences that could be delineated and that need to be considered in order to obtain meaningful results.

The first important finding is the need to account for unbalanced data in classification models (chapters 2, 3, 4, 7, 10) as the ratio between two response classes has considerable influence on the outcome. The first consequence was the utilization of a validation measure that is unaffected by class imbalances. Accuracy is one of the standard measures being used for validating classification models, but is not suitable for unbalanced data as the simple prediction of the majority class can lead to high accuracy (interpreted as perfect results) even though no delineation between the classes occurred. In this context, metrics that consider hits and misses per class are required. The Kappa index could be more appropriate and applicable for more than two-class problems, however, for two-class problems as applied in this thesis, ROC is the measure of choice as it is even independent of the probability threshold applied to delineate two classes. Even though ROC can provide an objective idea of the model's ability to delineate the classes, it does not solve the problem that algorithms tend to over-predict the majority class. The traditional approach to do this is by over- or under-sampling (see e.g. Kühnlein, 2014, at the example of rainfall retrievals). However, this thesis could show an objective, easy to operationalize and robust approach based on probability threshold tuning (see detailed description in chapter 2). The ROC-based distance to a perfect model as an alternative performance measure, allows identifying optimal probability thresholds for delineating two classes. When this measure is incorporated into model tuning, it is possible to fix the threshold in a robust way. This is especially important when the datasets are smaller (chapter 4 compared to chapter 2) as model performances of small datasets are more

sensitive to data splitting.

The size of a dataset is generally decisive for its vulnerability to data splitting and cross-validation strategies. For large datasets, a cross-validation is sufficient in order to get robust performance estimates as the variance in the fold-dependent performances are rather small (chapter 4). Small datasets have a higher variability according to data splitting and therefore CV needs to be repeated in order to get robust error estimates (chapter 6). It is certainly advisable for large datasets to leave complete parts of the data out (e.g. the data of an entire year as in chapter 4), which is usually not possible for small datasets (chapter 6). Also large space-time datasets which have a large number of data points in total, but sampled from a limited number of sampling locations are advised to rely on CV to get robust error estimates without "wasting" sampling locations for validation (chapters 5, 9).

It might be conspicuous in this thesis, that Random Forest was most frequently used as machine learning algorithm (chapters 6, 7, 8, 9, 10). In the course of the thesis, I often came across the question of why a certain algorithm was used. In the case of Random Forest, there are several reasons. Firstly, the underlying concept of regression or decision trees allows the inclusion of all types of variables without the need for normalisation procedures. If data need to be normalised, this becomes problematic when the model is later applied to new data as scaling requires the new data to be handled in consideration of scaling parameters of the input data. Though this is possible (chapters 2 - 4), it might generate a potential error source as it requires a constant awareness of this process. A second advantage of Random Forest is the robustness in view to a tuning of the hyperparameters. In contrast to many other well-known algorithms, Random Forest only features one tuning parameter (mtry) that is comparably insensitive, especially when only a small number of predictor variables is used. However, Random Forest is not necessarily the optimal algorithm for each dataset. Therefore, during a machine learning application it is always a question of which algorithm to use. In this thesis, two chapters extensively compare machine learning algorithms to find the optimal algorithm for different applications (chapters 2 and 5). As expected, the findings show that there is not a unique solution as the optimal algorithm is dependent on the dataset. However, especially the large rainfall dataset (chapter 2) showed robust results independently of the algorithm being used. In this context it was computation time that was decisive for choosing neural networks as the algorithm that was finally applied. The smaller and more variable air temperature dataset (chapter 5) showed a greater vulnerability to the choice of the algorithm. This highlights the need for testing different algorithms especially when the dataset is small. Finally, when choosing the algorithm, one must also consider the aim of the study. If not only the prediction is in the foreground but also an explicit understanding of the relative importance of the driving forces, an

algorithm must be chosen that can, despite the black-box concept, provide comparably reliable estimates of the importances (see Random Forest for identifying drivers of Rockfall, chapter 10).

Finally, I would like to highlight once again the necessity of the target-oriented validation strategy that is the superior pre-condition for a successful model application (chapter 11, section 12.4). A geographic machine learning application (especially if it is spatio-temporal) must include an appropriate validation strategy and a test for over-fitting variables.

12.6 Outlook

Miller and Goodchild (2015) describe data-driven geography being "evolutionary" rather than "revolutionary". Under this perspective it is understandable that applications of machine learning in geography still require ongoing refining on their way becoming a mature tool for environmental geography. Today, machine learning applications flood environmental science and its application is seen as being a great innovation in geography. However, it has to be clearly stated that for the next phase of successful evolution, one must step back from the idea of an uncontested tool and focus on its essential elements to make substantial contributions to environmental sciences. Applications must focus on objective studies of reliability and might consider the most simple method rather than the most complex one. New software implementation and refinements are needed that are not only focussing on making machine learning easy to use but more importantly help to get grip on objective modelling and validation strategies. This is especially important in view to spatial and spatio-temporal data as this is a field with special requirements. While this is an appeal to pause and carefully refine and adjust existing machine learning strategies to geographic data, the evolution of machine learning in geography did certainly not yet reach its climax. Especially for large geographic data, deep learning applications might be able to make further sense of unstructured data thus might have high potential especially for remote sensing.

12.7 Concluding remarks

This thesis showed the opportunities of machine learning for various applications in a wide range of environmental geography. It was emphasized that machine learning offers great opportunities yet challenges arise from the characteristics of geographic data. In order to produce valuable results, high expertise of the methods as well as of the scientific fields is the precondition to advance

the field of environmental geography using machine learning. In this context, a variety of environmental monitoring products was developed in this thesis that provide important baselines for the analysis of spatio-temporal processes. From a methodological perspective, the thesis raises, for the first time, the awareness of spatial or spatio-temporal over-fitting in geographic machine learning applications and the significant consequences on the outcome. In view to a solution, the thesis showed how a newly designed forward feature selection in conjunction with target-oriented validation strategies can be used to detect and avoid this problem and lead to objective and significant results. This thesis is also an appeal to think beyond the "standard way" of machine learning as the comparably easy accessibility of these methods nowadays leads to the risk of "blind" model applications. Environmental modelling is not simply about creating a map, but we must consider objective, target-oriented and robust modelling and validation strategies. Only when this is considered, machine learning provides a powerful tool to create scientifically valuable results for environmental geography.

References

- Ackerman, S. A., K. I. Strabala, W. P. Menzel, R. A. Frey, C. C. Moeller, and L. E. Gumley, 1998: Discriminating clear sky from clouds with MODIS. *Journal of Geophysical Research: Atmospheres*, **103** (D24), 32 141–32 157.
- Adler, R. F. and A. J. Negri, 1988: A Satellite Infrared Technique to Estimate Tropical Convective and Stratiform Rainfall. *Journal of Applied Meteorology*, **27**, 30–51.
- Aguilera, A. M., M. Escabias, and M. J. Valderrama, 2006: Using principal components for estimating logistic regression with high-dimensional multicollinear data. *Computational Statistics & Data Analysis*, **50** (8), 1905–1924.
- Aher, M., S. Pradhan, and Y. Dandawate, 2014: Rainfall Estimation over Roof-Top Using Land-Cover Classification of Google Earth Images. *International Conference on Electronic Systems, Signal Processing and Computing Technologies (ICESC)*, 111–116.
- Ahmad, S., A. Kalra, and H. Stephen, 2010: Estimating soil moisture using remote sensing data: A machine learning approach. *Advances in Water Resources*, **33** (1), 69–80.
- Al-Mulla, Y., J. Wu, P. Singh, M. Flury, W. Schillinger, D. Huggins, and C. Stöckle, 2009: Soil water and temperature in chemical versus reduced-tillage fallow in a mediterranean climate. *Applied Engineering in Agriculture*, **25**, 45–54.
- Alef, K. and P. Nannipieri, (Eds.) , 1995: *Methods in Applied Soil Microbiology and Biochemistry*. Academic Press, London.
- Ali, I., F. Greifeneder, J. Stamenkovic, M. Neumann, and C. Notarnicola, 2015: Review of Machine Learning Approaches for Biomass and Soil Moisture Retrievals from Remote Sensing Data. *Remote Sensing*, **7** (12), 16 398–16 421.
- Allen, S. and C. Huggel, 2013: Extremely warm temperatures as a potential cause of recent high mountain rockfall. *Global and Planetary Change*, **107**, 59–69.
- Allen, S. K., S. Gruber, and I. F. Owens, 2009: Exploring steep bedrock permafrost and its relationship with recent slope failures in the Southern Alps of New Zealand. *Permafrost and Periglacial Processes*, **20** (4), 345–356.
- Allen Jr., R., P. A. Durkee, and C. H. Wash, 1990: Snow/Cloud Discrimination with Multispectral Satellite Measurements. *Journal of Applied Meteorology*, **29** (10), 994–1004.
- Allison, P., 2001: *Logistic Regression Using the SAS System: Theory and Application*. SAS Publishing, North Carolina.
- Almeer, M. H., 2012: Vegetation extraction from free Google Earth images of deserts using a robust BPNN approach in HSV space. *International Journal of Advanced Research in Computer and Communication Engineering*, **1**, 134–140.

- Ameur, Z., S. Ameur, A. Adane, H. Sauvageot, and K. Bara, 2004: Cloud classification using the textural features of Meteosat images. *International Journal of Remote Sensing*, **25** (21), 4491–4503.
- Aminou, D. M. A., B. Jacquet, and F. Pasternak, 1997: Characteristics of the Meteosat Second Generation (MSG) radiometer/imager: SEVIRI. *Proceedings of SPIE: Sensors, Systems, and Next-Generation Satellites*, 3221, 19–31, 3221.
- Amitrano, D., S. Gruber, and L. Girard, 2012: Evidence of frost-cracking inferred from acoustic emissions in a high-alpine rock-wall. *Earth and Planetary Science Letters*, **341–344**, 86–93.
- Anderson, C., 2008: The end of theory: The data deluge makes the scientific method obsolete. *Wired*, **16** (7).
- André, M., 2002: Rates of postglacial rock weathering on glacially scoured outcrops (Abisko–Riksgränsen area, 68°N). *Geografiska Annaler*, **A84**, 139–150.
- André, M.-F., 1997: Holocene Rockwall Retreat in Svalbard: A Triple-Rate Evolution. *Earth Surface Processes and Landforms*, **22** (5), 423–440.
- Appelhans, T., E. Mwangomo, D. R. Hardy, A. Hemp, and T. Nauss, 2015: Evaluating machine learning approaches for the interpolation of monthly air temperature at Mt. Kilimanjaro, Tanzania. *Spatial Statistics*, **14**, Part A, 91–113.
- Arkin, P. A. and B. N. Meisner, 1987: The Relationship between Large-Scale Convective Rainfall and Cold Cloud over the Western Hemisphere during 1982–84. *Monthly Weather Review*, **115** (1), 51–74.
- Asefa, T., M. Kemblowski, M. McKee, and A. Khalil, 2006: Multi-time scale stream flow predictions: The support vector machines approach. *Journal of Hydrology*, **318** (1–4), 7–16.
- Asner, G., A. Elmore, L. Olander, R. Martin, and A. Harris, 2004: Grazing systems, ecosystem responses, and global change. *Annual Review of Environment and Resources*, **29**, 261–299.
- Atkinson, P., H. Jiskoot, R. Massari, and T. Murray, 1998: Generalized linear modelling in geomorphology. *Earth Surface Processes and Landforms*, **23** (13), 1185–1195.
- Augustinus, P., 1995: Rock mass strength and the stability of some glacial valley slopes. *Zeitschrift für Geomorphologie*, **39**, 55–68.
- Augustinus, P. C., 1992: The influence of rock mass strength on glacial valley cross-profile morphometry: A case study from the Southern Alps, New Zealand. *Earth Surface Processes and Landforms*, **17** (1), 39–51.
- Ba, M. B. and A. Gruber, 2001: GOES Multispectral Rainfall Algorithm (GM-SRA). *Journal of Applied Meteorology*, **40**, 1500–1514.
- Babel, W., et al., 2014: Pasture degradation modifies the water and carbon cycles of the Tibetan highlands. *Biogeosciences*, **11**, 8861–8923.

- Badreldin, N. and A. Sanchez-Azofeifa, 2015: Estimating Forest Biomass Dynamics by Integrating Multi-Temporal Landsat Satellite Images with Ground and Airborne LiDAR Data in the Coal Valley Mine, Alberta, Canada. *Remote Sensing*, **7** (3), 2832–2849.
- Bai, S.-B., J. Wang, G.-N. Lü, P.-G. Zhou, S.-S. Hou, and S.-N. Xu, 2010: GIS-based logistic regression for landslide susceptibility mapping of the Zhongxian segment in the Three Gorges area, China. *Geomorphology*, **115** (1–2), 23–31.
- Baldeck, C. A., G. P. Asner, R. E. Martin, C. B. Anderson, D. E. Knapp, J. R. Kellner, and S. J. Wright, 2015: Operational Tree Species Mapping in a Diverse Tropical Forest with Airborne Imaging Spectroscopy. *PLoS ONE*, **10** (7), e0118403.
- Ballantyne, C. K., 2002: A general model of paraglacial landscape response. *The Holocene*, **12** (3), 371–376.
- Ballantyne, C. K. and J. O. Stone, 2013: Timing and periodicity of paraglacial rock-slope failures in the Scottish Highlands. *Geomorphology*, **186**, 150–161.
- Ballantyne, C. K., P. Wilson, D. Gheorghiu, and A. Rodés, 2014: Enhanced rock-slope failure following ice-sheet deglaciation: timing and causes. *Earth Surface Processes and Landforms*, **39** (7), 900–913.
- Baltensperger, A. P. and F. Huettmann, 2015: Predictive spatial niche and biodiversity hotspot models for small mammal communities in Alaska: applying machine-learning to conservation planning. *Landscape Ecology*, **30** (4), 681–697.
- Bárdossy, A. and W. Lehmann, 1998: Spatial distribution of soil moisture in a small catchment. Part 1: geostatistical analysis. *Journal of Hydrology*, **206** (1), 1–15.
- Barnes, E., T. Clarke, S. Richards, P. Colaizzi, J. Haberland, M. Kostrzewski, P. Waller, C. Choi, E. Riley, T. Thompson, R. Lascano, H. Li, and M. Moran, (Eds.), 2000: *Proceedings of the international conference of precision agriculture*, Vol. 5, Bloomington, MN, 16–19 July.
- Bartels, H., E. Weigl, T. Reich, P. Lang, A. Wagner, O. Kohler, N. Gerlach, and M. GmbH, 2004: *Projekt RADOLAN - Routineverfahren zur Online-Aneichung der Radarniederschlagsdaten mit Hilfe von automatischen Bodenniederschlagsstationen (Ombrometer)*. Deutscher Wetterdienst, Offenbach, 1–111 pp.
- Bearth, P., 1980: *Geologischer Atlas Schweiz 1:25000. Blatt: 1308 St. Niklaus. Erläuterungen*. Kümmerly & Frey AG, Bern.
- Beaudette, D. and P. Roudier, 2013: *aqp: Algorithms for Quantitative Pedology*. URL <http://CRAN.R-project.org/package=aqp>, R package version 1.4.
- Beerli, O., R. Phillips, J. Hendrickson, A. B. Frank, and S. Kronberg, 2007: Estimating forage quantity and quality using aerial hyperspectral imagery for northern mixed-grass prairie. *Remote Sensing of Environment*, **110** (2), 216–225.

- Behrangi, A., K. Hsu, B. Imam, S. Sorooshian, G. Huffman, and R. Kuligowski, 2009a: PERSIANN-MSA: A Precipitation Estimation Method from Satellite-Based Multispectral Analysis. *Journal of Hydrometeorology*, **10** (6), 1414–1429.
- Behrangi, A., K. Hsu, B. Imam, S. Sorooshian, and R. Kuligowski, 2009b: Evaluating the Utility of Multispectral Information in Delineating the Areal Extent of Precipitation. *Journal of Hydrometeorology*, **10** (3), 684–700.
- Benali, A., A. Carvalho, J. Nunes, N. Carvalhais, and A. Santos, 2012: Estimating air surface temperature in Portugal using MODIS LST data. *Remote Sensing of Environment*, **124**, 108–121.
- Benas, N., S. Finkensieper, M. Stengel, G.-J. van Zadelhoff, T. Hanschmann, R. Hollmann, and J. F. Meirink, 2017: The MSG-SEVIRI based cloud property data record CLAAS-2. *Earth System Science Data Discussions*, **2017**, 1–36.
- Beven, K. J. and M. J. Kirkby, 1979: A physically based, variable contributing area model of basin hydrology / Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. *Hydrological Sciences Bulletin*, **24** (1), 43–69.
- Bivand, R., T. Keitt, and B. Rowlingson, 2014: *rgdal: Bindings for the Geospatial Data Abstraction Library*. URL <https://CRAN.R-project.org/package=rgdal>, R package version 0.9-1.
- Blagodatsky, S. A., O. Heinemeyer, and J. Richter, 2000: Estimating the active and total soil microbial biomass by kinetic respiration analysis. *Biology and Fertility of Soils*, **32** (1), 73–81.
- Böhme, M., M.-H. Derron, and M. Jaboyedoff, 2014: Quantitative spatial analysis of rockfalls from road inventories: a combined statistical and physical susceptibility model. *Natural Hazards and Earth System Sciences Discussions*, **2**, 81–121.
- Bonan, G., 2008: *Ecological climatology. concepts and applications*. 2d ed., Cambridge University Press, Cambridge.
- Borgomeo, E., K. V. Hebdict, A. C. Whittaker, and L. Lonergan, 2014: Characterising the spatial distribution, frequency and geomorphic controls on landslide occurrence, Molise, Italy. *Geomorphology*, **226**, 148–161.
- Bornemann, L., G. Welp, and W. Amelung, 2010: Particulate organic matter at the field scale: Rapid acquisition using mid-infrared spectroscopy. *Soil Science Society of America Journal*, **74**, 1147–1156.
- Bornemann, L., G. Welp, S. Brodowski, A. Rodionov, and W. Amelung, 2008: Rapid assessment of black carbon in soil organic matter using mid-infrared spectroscopy. *Organic Geochemistry*, **39** (11), 1537–1544.
- Bosatta, E. and G. Agren, 1999: Soil organic matter quality interpreted thermodynamically. *Soil Biology and Biochemistry*, **31** (13), 1889–1891.
- Bossel, H., 1994: *Modeling and Simulation*. Springer.

- Bovis, M. J., 1990: Rock-slope deformation at Affliction Creek, southern Coast Mountains, British Columbia. *Canadian Journal of Earth Sciences*, **27** (2), 243–254.
- Bowden, R. D., K. M. Newkirk, and G. M. Rullo, 1998: Carbon dioxide and methane fluxes by a forest soil under laboratory-controlled moisture and temperature conditions. *Soil Biology and Biochemistry*, **30** (12), 1591–1597.
- Bradley, A. P., 1997: The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, **30** (7), 1145–1159.
- Breiman, L., 2001a: Random Forests. *Machine Learning*, **45** (1), 5–32.
- Breiman, L., 2001b: Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). *Statistical Science*, **16** (3), 199–231.
- Brenning, A., 2005: Spatial prediction models for landslide hazards: review, comparison and evaluation. *Natural Hazards and Earth System Sciences*, **5** (6), 853–862.
- Brenning, A., 2013: *RSAGA: SAGA Geoprocessing and Terrain Analysis in R*. URL <http://CRAN.R-project.org/package=RSAGA>, R package version 0.93-6.
- Brenning, A., S. Long, and P. Fieguth, 2012: Detecting rock glacier flow structures using Gabor filters and IKONOS imagery. *Remote Sensing of Environment*, **125**, 227–237.
- Breulmann, M., N. P. Masyutenko, B. M. Kogut, R. Schroll, U. Dörfler, F. Buscot, and E. Schulz, 2014: Short-term bioavailability of carbon in soil organic matter fractions of different particle sizes and densities in grassland ecosystems. *Science of The Total Environment*, **497**, 29–37.
- Brevik, E. C., C. Calzolari, B. A. Miller, P. Pereira, C. Kabala, A. Baumgarten, and A. Jordán, 2016: Soil mapping, classification, and pedologic modeling: History and future directions. *Geoderma*, **264**, Part B, 256–274.
- Britz, M.-L. and D. Ward, 2007: Dynamics of woody vegetation in a semi-arid savanna, with a focus on bush encroachment. *African Journal of Range and Forage Science*, **24** (3), 131–140.
- Bromwich, D. H., J. P. Nicolas, A. J. Monaghan, M. A. Lazzara, L. M. Keller, G. A. Weidner, and A. B. Wilson, 2013: Central West Antarctica among the most rapidly warming regions on Earth. *Nature Geoscience*, **6** (2), 139–145.
- Brown, D. G., D. P. Lusch, and K. A. Duda, 1998: Supervised classification of types of glaciated landscapes using digital elevation data. *Geomorphology*, **21** (3), 233–250.
- Brungard, C. W., J. L. Boettinger, M. C. Duniway, S. A. Wills, and T. C. E. Jr., 2015: Machine learning for predicting soil classes in three semi-arid landscapes. *Geoderma*, **239–240**, 68–83.
- Brunner, F. K. and A. E. Scheidegger, 1973: Exfoliation. *Rock mechanics*, **5** (1), 43–62.

- Brunsdon, D. and R. H. Kesel, 1973: Slope Development on a Mississippi River Bluff in Historic Time. *The Journal of Geology*, **81** (5), 576–598.
- Bruzzzone, L. and F. Melgani, 2005: Robust multiple estimator systems for the analysis of biophysical parameters from remotely sensed data. *IEEE Transactions on Geoscience and Remote Sensing*, **43** (1), 159–174.
- Buitenwerf, R., W. J. Bond, N. Stevens, and W. S. W. Trollope, 2012: Increased tree densities in South African savannas: < 50 years of data suggests CO₂ as a driver. *Global Change Biology*, **18** (2), 675–684.
- Burrough, P., (Ed.) , 1998: *Principles of Geographical Information Systems*. 2d ed., Oxford University Press, Oxford.
- Büdel, J., (Ed.) , 1977: *Climatic Geomorphology*. Princeton University Press, Princeton, N.J.
- Caine, N., 1974: The geomorphic processes of the alpine environment. *Arctic and Alpine Environments*, J. D. Ives and R. G. Barry, Eds., Methuen, London, 721–748.
- Cammeraat, E., R. van Beek, and A. Kooijman, 2005: Vegetation Succession and its Consequences for Slope Stability in SE Spain. *Plant and Soil*, **278** (1), 135–147.
- Camps-Valls, G., 2009: Machine learning in remote sensing data processing. *IEEE International Workshop on Machine Learning for Signal Processing*, 1–6.
- Cantón, Y., A. Solé-Benet, and F. Domingo, 2004: Temporal and spatial patterns of soil moisture in semiarid badlands of SE Spain. *Journal of Hydrology*, **285** (1–4), 199–214.
- Cao, J. J., E. T. Yeh, N. M. Holden, Y. Y. Yang, and G. Z. Du, 2013: The effects of enclosures and land-use contracts on rangeland degradation on the Qinghai-Tibetan Plateau. *Journal of Arid Environments*, **97**, 3–8.
- Capacci, D. and B. J. Conway, 2005: Delineation of precipitation areas from MODIS visible and infrared imagery with artificial neural networks. *Meteorological Applications*, **12** (4), 291–305.
- Caputo, B., K. Sim, F. Furesjo, and A. Smola, 2002: Appearance-based object recognition using SVMs: which kernel should I use? *Proc of NIPS workshop on Statistical methods for computational experiments in visual processing and computer vision*.
- Carr, D., 2014: *hexbin: Hexagonal Binning Routines*. URL <http://CRAN.R-project.org/package=hexbin>, R package version 1.26-2.
- Carrara, A., F. Guzzetti, M. Cardinali, and P. Reichenbach, 1999: Use of GIS Technology in the Prediction and Monitoring of Landslide Hazard. *Natural Hazards*, **20** (2), 117–135.
- Carson, M. A. and D. J. Petley, 1970: The Existence of Threshold Hillslopes in the Denudation of the Landscape. *Transactions of the Institute of British Geographers*, (49), 71–95.

- Carter, G. A. and A. K. Knapp, 2001: Leaf optical properties in higher plants: linking spectral characteristics to stress and chlorophyll concentration. *American Journal of Botany*, **88** (4), 677–684.
- Catani, F., D. Lagomarsino, S. Segoni, and V. Tofani, 2013: Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues. *Natural Hazards and Earth System Sciences*, **13** (11), 2815–2831.
- Cattani, E., A. Merino, and V. Levizzani, 2016: Evaluation of Monthly Satellite-Derived Precipitation Products over East Africa. *Journal of Hydrometeorology*, **17** (10), 2555–2573.
- Ceccato, P., K. Fernandes, D. Ruiz, and E. Allis, 2014: Climate and environmental monitoring for decision making. *Earth Perspectives*, **1** (1), 16.
- Černohlávková, J., J. Jarkovský, M. Nešporová, and J. Hofman, 2009: Variability of soil microbial properties: Effects of sampling, handling and storage. *Ecotoxicology and Environmental Safety*, **72** (8), 2102–2108.
- Chang, C.-W., D. Laird, M. Mausbach, and C. R. Hurburgh Jr., 2001: Near-Infrared Reflectance Spectroscopy–Principal Components Regression Analyses of Soil Properties. *Soil Science Society of America Journal* 5:480–490. *Soil Science Society of America Journal*, **65** (2), 480–490.
- Chen, J., H.-P. Blume, and L. Beyer, 2000: Weathering of rocks induced by lichen colonization – a review. *Catena*, **39** (2), 121–146.
- Christodoulou, C. I., S. C. Michaelides, and C. S. Pattichis, 2003: Multifeature Texture Analysis for the Classification of Clouds in Satellite Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, **41**.
- Church, M., 1996: Space, time and the mountain: how do we order what we see? *The Scientific Nature of Geomorphology*, B. Rhoads and C. Thorn, Eds., Wiley, Chichester, 147–170.
- Church, M. and J. Ryder, 1972: Paraglacial Sedimentation: A Consideration of Fluvial Processes Conditioned by Glaciation. *Geological Society of America Bulletin*, **83** (10), 3059–3072.
- Cobo, J. G., G. Dercon, T. Yekeye, L. Chapungu, C. Kadzere, A. Murwira, R. Delve, and G. Cadisch, 2010: Integration of mid-infrared spectroscopy and geostatistics in the assessment of soil spatial variability at landscape level. *Geoderma*, **158** (3), 398–411.
- Coe, J. A. and E. L. Harp, 2007: Influence of tectonic folding on rockfall susceptibility, American Fork Canyon, Utah, USA. *Natural Hazards and Earth System Sciences*, **7** (1), 1–14.
- Colombi, A., C. De Michele, M. Pepe, and A. Rampini, 2007: Estimation of daily mean air temperature from MODIS LST in alpine areas. *EARSeL eProceedings*, Vol. 6.
- Conant, R. T., R. A. Drijber, M. L. Haddix, W. J. Parton, E. A. Paul, A. F. Plante, J. Six, and J. M. Steinweg, 2008: Sensitivity of organic matter decomposition to warming varies with its quality. *Global Change Biology*, **14** (4), 868–877.

- Conrad, O., B. Bechtel, M. Bock, H. Dietrich, E. Fischer, L. Gerlitz, J. Wehberg, V. Wichmann, and J. Böhner, 2015: System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geoscientific Model Development*, **8** (7), 1991–2007.
- Convey, P., *et al.*, 2014: The spatial structure of Antarctic biodiversity. *Ecological Monographs*, **84** (2), 203–244.
- Corwin, D. and S. Lesch, 2005: Apparent soil electrical conductivity measurements in agriculture. *Computers and Electronics in Agriculture*, **46** (1–3), 11–43.
- Cossart, E., D. Mercier, A. Decaulne, T. Feuillet, H. P. Jónsson, and p. Saeed Emamifarmundsson, 2014: Impacts of post-glacial rebound on landslide spatial distribution at a regional scale in northern Iceland (Skagafjörður). *Earth Surface Processes and Landforms*, **39** (3), 336–350.
- Cossart, E. E., 2008: Landform connectivity and waves of negative feedbacks during the paraglacial period, a case study : the Tabuc subcatchment since the end of the Little Ice Age (massif des Écrins, France). *Géomorphologie : relief, processus, environnement*, **4**, 249–260.
- Coutard, J.-P. and B. Francou, 1989: Rock Temperature Measurements in Two Alpine Environments: Implications for Frost Shattering. *Arctic and Alpine Research*, **21** (4), 399–416.
- Cozzolino, D., F. Montossi, and R. San Julian, 2005: The use of visible (VIS) and near infrared (NIR) reflectance spectroscopy to predict fibre diameter in both clean and greasy wool samples. *Animal Science*, **80** (3), 333–337.
- Craine, J. M., N. Fierer, and K. K. McLauchlan, 2010: Widespread coupling between the rate and temperature sensitivity of organic matter decay. *Nature Geoscience*, **3** (12), 854–857.
- Craine, J. M. and T. M. Gelderman, 2011: Soil moisture controls on temperature sensitivity of soil organic carbon decomposition for a mesic grassland. *Soil Biology and Biochemistry*, **43** (2), 455–457.
- Crozier, M., (Ed.) , 1989: *Landslides: Causes, Consequences and Environment*. Routledge, London.
- Cruden, D. M. and X.-Q. Hu, 1994: Topples on underdip slopes in the Highwood Pass, Alberta, Canada. *Quarterly Journal of Engineering Geology and Hydrogeology*, **27** (1), 57–68.
- Cruden, D. M. and X. Q. Hu, 1998: *Landslides in the Rocky Mountains of Canada*, 133–148. Springer Netherlands, Dordrecht.
- Curatola Fernández, G. F., W. A. Obermeier, A. Gerique, M. F. L. Sandoval, L. W. Lehnert, B. Thies, and J. Bendix, 2015: Land Cover Change in the Andes of Southern Ecuador - Patterns and Drivers. *Remote Sensing*, **7** (3), 2509–2542.
- Curry, A. M. and C. J. Morris, 2004: Lateglacial and Holocene talus slope development and rockwall retreat on Mynydd Du, UK. *Geomorphology*, **58** (1–4), 85–106.

- Cutler, D. R., T. C. Edwards, K. H. Beard, A. Cutler, K. T. Hess, J. Gibson, and J. J. Lawler, 2007: Random Forests for Classification in Ecology. *Ecology*, **88** (11), 2783–2792.
- Davies, M. C. R., O. Hamza, and C. Harris, 2001: The effect of rise in mean annual temperature on the stability of rock slopes containing ice-filled discontinuities. *Permafrost and Periglacial Processes*, **12** (1), 137–144.
- de Boer, D. H., 1992: Hierarchies and spatial scale in process geomorphology: a review. *Geomorphology*, **4** (5), 303–318.
- De Neve, S., J. Van de Steene, R. Hartmann, and G. Hofman, 2000: Using time domain reflectometry for monitoring mineralization of nitrogen from soil organic matter. *European Journal of Soil Science*, **51** (2), 295–304.
- Decagon Devices, Inc., 2014: *5TE Water Content, EC and Temperature Sensor*. Pullman, WA. URL http://manuals.decagon.com/Manuals/13509_5TE_Web.pdf.
- Diggle, P. and P. J. Ribeiro, 2007: *Model-based Geostatistics*. Springer Series in Statistics, Springer.
- Dixon, J. C. and C. E. Thorn, 2005: Chemical weathering and landscape development in mid-latitude alpine environments. *Geomorphology*, **67** (1), 127–145.
- Doran, P. T., G. Dana, J. T. Hastings, and R. A. Wharton, 1995: McMurdo Dry Valleys Long-Term Ecological Research (LTER): LTER automatic weather network (LAWN). *Antarctic Journal of the U.S.*, **30** (5), 276–280.
- Dorren, L. K. and A. C. Seijmonsbergen, 2003: Comparison of three GIS-based models for predicting rockfall runout zones at a regional scale. *Geomorphology*, **56** (1–2), 49–64.
- Draebing, D., M. Krautblatter, and R. Dikau, 2014: Interaction of thermal and mechanical processes in steep permafrost rock walls: A conceptual approach. *Geomorphology*, **226**, 226–235.
- Dreber, N., C. J. Harmse, A. Götze, W. S. W. Trollope, and K. Kellner, 2014: Quantifying the woody component of savanna vegetation along a density gradient in the Kalahari Bushveld: a comparison of two adapted point-centered quarter methods. *Rangeland Journal*, **36** (1), 91–103.
- Duarte, R. and J. Marquinez, 2002: The influence of environmental and lithologic factors on rockfall at a regional scale: an evaluation using GIS. *Geomorphology*, **43** (1–2), 117–136.
- D’Acqui, L. P., A. Pucci, and L. J. Janik, 2010: Soil properties prediction of western Mediterranean islands with similar climatic environments by means of mid-infrared diffuse reflectance spectroscopy. *European Journal of Soil Science*, **61** (6), 865–876.
- Eigenberg, R., J. Doran, J. Nienaber, R. Ferguson, and B. Woodbury, 2002: Electrical conductivity monitoring of soil condition and available N with animal manure and a cover crop. *Agriculture, Ecosystems & Environment*, **88** (2), 183–193.

- Eitel, J., D. Long, P. Gessler, E. Hunt, and D. Brown, 2009: Sensitivity of ground-based remote sensing estimates of wheat chlorophyll content to variation in soil reflectance. *Soil Science Society of America*, **73**, 1715–1723.
- Eitel, J., D. Long, P. Gessler, and A. Smith, 2007: Using in-situ measurements to evaluate the new rapideye satellite series for prediction of wheat nitrogen status. *International Journal of Remote Sensing*, **28** (18), 4183–4190.
- Eitel, J. U., L. A. Vierling, M. E. Litvak, D. S. Long, U. Schulthess, A. A. Ager, D. J. Krofcheck, and L. Stoscheck, 2011: Broadband, red-edge information from satellites improves early stress detection in a New Mexico conifer woodland. *Remote Sensing of Environment*, **115** (12), 3640–3646.
- Eitel, J. U. H., D. S. Long, P. E. Gessler, and E. R. Hunt, 2008: Combined Spectral Index to Improve Ground-Based Estimates of Nitrogen Status in Dryland Wheat. *Agronomy*, **100** (6), 1694–1702.
- Emamifar, S., A. Rahimikhoob, and A. A. Noroozi, 2013: Daily mean air temperature estimation from MODIS land surface temperature products based on M5 model tree. *International Journal of Climatology*, **33** (15), 3174–3181.
- Escabias, M., A. M. Aguilera, and M. J. Valderrama, 2005: Modeling environmental data by functional principal component logistic regression. *Environmetrics*, **16** (1), 95–107.
- Escher, H., 1970: Die Bestimmung der klimatischen Schneegrenze in den Schweizer Alpen. *Geographica Helvetica*, **25** (1), 35–43.
- ESRI, (Ed.) , 2006: *ArcGIS 10.1*. Environmental System Research Institute, Inc., Redlands, CA.
- EUMETSAT, 2010: High Rate SEVIRI Level 1.5 Image Data - MSG - 0 degree. URL <http://navigator.eumetsat.int/discovery/Start/DirectSearch/DetailResult.do?f%28r0%29=E0:EUM:DAT:MSG:HRSEVIRI>, (last access: 13 July 2015).
- EUMETSAT, 2012a: *Conversion from radiances to reflectances for SEVIRI warm channels*.
- EUMETSAT, 2012b: *The Conversion from Effective Radiances to Equivalent Brightness Temperatures*.
- Evans, S. and O. Hungr, 1993: The assessment of rockfall hazard at the base of talus slopes. *Canadian Geotechnical Journal*, **30** (4), 620–636.
- Evans, S. G. and J. J. Clague, 1994: Recent climatic change and catastrophic geomorphic processes in mountain environments. *Geomorphology*, **10** (1), 107–128.
- Fava, F., G. Parolo, R. Colombo, F. Gusmeroli, G. Della Marianna, A. Monteiro, and S. Bocchi, 2010: Fine-scale assessment of hay meadow productivity and plant diversity in the European Alps using field spectrometric data. *Agriculture, Ecosystems & Environment*, **137**, 151–157.
- Fawcett, T., 2006: An introduction to ROC analysis. *Pattern Recognition Letters*, **27**, 861–874.

- Feidas, H. and A. Giannakos, 2012: Classifying convective and stratiform rain using multispectral infrared Meteosat Second Generation satellite data. *Theoretical and Applied Climatology*, **108** (3–4), 613–630.
- Ferrier, S., 2002: Mapping Spatial Pattern in Biodiversity for Regional Conservation Planning: Where to from Here? *Systematic Biology*, **51** (2), 331–363.
- Feuillet, T., J. Coquin, D. Mercier, E. Cossart, A. Decaulne, H. P. Jónsson, and Þorsteinn Sæmundsson, 2014: Focusing on the spatial non-stationarity of landslide predisposing factors in northern Iceland. *Progress in Physical Geography*, **38** (3), 354–377.
- Fierer, N., B. P. Colman, J. P. Schimel, and R. B. Jackson, 2006: Predicting the temperature dependence of microbial respiration in soil: A continental-scale analysis. *Global Biogeochemical Cycles*, **20** (3).
- Finkensieper, S., J.-F. Meirink, G.-J. van Zadelhoff, T. Hanschmann, N. Benas, M. Stengel, P. Fuchs, R. Hollmann, and M. Werscheck, 2016: CLAAS-2: CM SAF CLOUD property dAtAset using SEVIRI - Edition 2. Tech. rep., Satellite Application Facility on Climate Monitoring.
- Fischer, L., 2010: *Slope instabilities on perennially frozen and glacierized rock walls: multi-scale observations, analysis and modelling PhD thesis Schriftenreihe Physische Geographie: Glaziologie und Geomorphodynamik 58. University of Zürich, Switzerland (University of Zurich, Zürich, 184 pp.)*, PhD thesis Schriftenreihe Physische Geographie: Glaziologie und Geomorphodynamik, Vol. 58. University of Zurich, Zürich.
- Fischer, L., A. Käab, C. Huggel, and J. Noetzli, 2006: Geology, glacier retreat and permafrost degradation as controlling factors of slope instabilities in a high-mountain rock wall: the Monte Rosa east face. *Natural Hazards and Earth System Sciences*, **6** (5), 761–772.
- Fischer, L., R. S. Purves, C. Huggel, J. Noetzli, and W. Haeberli, 2012: On the influence of topographic, geological and cryospheric factors on rock avalanches and rockfalls in high-mountain areas. *Natural Hazards and Earth System Sciences*, **12** (1), 241–254.
- Foereid, B., D. S. Ward, N. Mahowald, E. Paterson, and J. Lehmann, 2014: The sensitivity of carbon turnover in the Community Land Model to modified assumptions about soil processes. *Earth System Dynamics*, **5**, 211–221.
- Frattini, P., G. Crosta, A. Carrara, and F. Agliardi, 2008: Assessment of rock-fall susceptibility by integrating statistical and physically-based approaches. *Geomorphology*, **94** (3–4), 419–437.
- Fretwell, P., *et al.*, 2013: Bedmap2: improved ice bed, surface and thickness datasets for Antarctica. *The Cryosphere*, **7** (1), 375–393.
- Friedman, S. P., 2005: Soil properties influencing apparent electrical conductivity: a review. *Computers and Electronics in Agriculture*, **46** (1–3), 45–70.
- Fryirs, K., G. J. Brierley, and W. D. Erskine, 2012: Use of ergodic reasoning to reconstruct the historical range of variability and evolutionary trajectory of rivers. *Earth Surface Processes and Landforms*, **37** (7), 763–773.

- Fu, P. and P. M. Rich, 2002: A geometric solar radiation model with applications in agriculture and forestry. *Computers and Electronics in Agriculture*, **37** (1–3), 25–35.
- Fynn, R. and T. O'Connor, 2000: Effect of stocking rate and rainfall on rangeland dynamics and cattle performance in a semi-arid savanna, South Africa. *Journal of Applied Ecology*, **37** (3), 491–507.
- Gallo, K., R. Hale, D. Tarpley, and Y. Yu, 2011: Evaluation of the Relationship between Air and Land Surface Temperature under Clear- and Cloudy-Sky Conditions. *Journal of Applied Meteorology and Climatology*, **50** (3), 767–775.
- Gao, Q. Z., Y. F. Wan, H. M. Xu, Y. Li, W. Z. Jiangcun, and A. Borjigidai, 2010: Alpine grassland degradation index and its response to recent climate variability in Northern Tibet, China. *Quaternary International*, **226** (1–2), 143–150.
- Garzón, M. B., R. Blazek, M. Neteler, R. S. de Dios, H. S. Ollero, and C. Furlanello, 2006: Predicting habitat suitability with machine learning models: The potential area of *Pinus sylvestris* L. in the Iberian Peninsula. *Ecological Modelling*, **197** (3–4), 383–393.
- Gasch, C. K., T. Hengl, B. Gräler, H. Meyer, T. S. Magney, and D. J. Brown, 2015: Spatio-temporal interpolation of soil water, temperature, and electrical conductivity in 3D + T: The Cook Agronomy Farm data set. *Spatial Statistics*, **14**, Part A, 70–90.
- Gerber, E., 1974: Klassifikation von Schutthalden. *Geographica Helvetica*, **29** (2–3), 73–82.
- Gerber, E. and A. Scheidegger, 1969: Stress-induced weathering of rock masses. *Eclogae Geologicae Helvetiae*, **62**, 401–416.
- Gessner, U., M. Machwitz, C. Conrad, and S. Dech, 2013: Estimating the fractional cover of growth forms and bare surface in savannas. A multi-resolution approach based on regression tree ensembles. *Remote Sensing of Environment*, **129**, 90–102.
- Gevrey, M., I. Dimopoulos, and S. Lek, 2003: Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological Modelling*, **160** (3), 249–264.
- Ghosh, A. and P. Joshi, 2014: A comparison of selected classification algorithms for mapping bamboo patches in lower Gangetic plains using very high resolution WorldView 2 imagery. *International Journal of Applied Earth Observation and Geoinformation*, **26**, 298–311.
- Giannakos, A. and H. Feidas, 2011: Detection of rainy clouds based on their spectral and textural features on Meteosat multispectral infrared data. *EU-METSAT Meteorological Satellite Conference*, Oslo, Norway.
- Giannakos, A. and H. Feidas, 2012: Precipitation estimation based on spectral and textural features of Meteosat multispectral infrared data. *EUMETSAT Meteorological Satellite Conference*, Sopot, Poland.

- Giannakos, A. and H. Feidas, 2013: Classification of convective and stratiform rain based on the spectral and textural features of Meteosat Second Generation infrared data. *Theoretical and Applied Climatology*, **113** (3-4), 495–510.
- Girard, L., S. Gruber, S. Weber, and J. Beutel, 2013: Environmental controls of frost cracking revealed through in situ acoustic emission measurements in steep bedrock. *Geophysical Research Letters*, **40** (9), 1748–1753.
- Gislason, P. O., J. A. Benediktsson, and J. R. Sveinsson, 2006: Random Forests for land cover classification. *Pattern Recognition Letters*, **27** (4), 294–300.
- Gizachew, B., S. Solberg, E. Næsset, T. Gobakken, O. M. Bollandsås, J. Breidenbach, E. Zahabu, and E. W. Mauya, 2016: Mapping and estimating the total living biomass and carbon in low-biomass woodlands using Landsat 8 CDR data. *Carbon Balance and Management*, **11**, 13.
- Global Carex Group, 2015: Making *Carex* monophyletic (Cyperaceae, tribe Cariceae): a new broader circumscription. *Botanical Journal of the Linnean Society*, **179** (1), 1–42.
- Goetz, J., A. Brenning, H. Petschko, and P. Leopold, 2015: Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. *Computers & Geosciences*, **81**, 1–11.
- Gokaraju, B., S. S. Durbha, R. L. King, and N. H. Younan, 2011: A Machine Learning Based Spatio-Temporal Data Mining Approach for Detection of Harmful Algal Blooms in the Gulf of Mexico. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **4** (3), 710–720.
- Gong Li, S., Y. Harazono, T. Oikawa, H. L. Zhao, Z. Ying He, and X. L. Chang, 2000: Grassland desertification by grazing and the resulting micrometeorological changes in Inner Mongolia. *Agricultural and Forest Meteorology*, **102** (2-3), 125–137.
- Goovaerts, P., 2000: Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of Hydrology*, **228** (1–2), 113–129.
- Govindaraju, R. and A. Rao, (Eds.) , 2000: *Artificial Neural Networks in Hydrology*. Springer Netherlands.
- Graham, M. H., 2003: Confronting Multicollinearity in Ecological Multiple Regression. *Ecology*, **84** (11), 2809–2815.
- Gray, W. and G. Gray, 2017: *Introduction to Environmental Modeling*. Cambridge University Press.
- Grimes, D., E. Coppola, M. Verdecchia, and G. Visconti, 2003: A neural network approach to real-time rainfall estimation for Africa using satellite data. *Journal of Hydrometeorology*, **4**, 1119–1133.
- Gritsch, C., M. Zimmermann, and S. Zechmeister-Boltenstern, 2015: Interdependencies between temperature and moisture sensitivities of CO₂ emissions in European land ecosystems. *Biogeosciences*, **12** (20), 5981–5993.
- Gruber, S., 2005: Mountain Permafrost: Transient Spatial Modelling, Model Verification and the Use of Remote Sensing. Ph.D. thesis, University of Zurich, Switzerland.

- Gruber, S. and W. Haeberli, 2007: Permafrost in steep bedrock slopes and its temperature-related destabilization following climate change. *Journal of Geophysical Research: Earth Surface*, **112** (F2).
- Gudmundsson, G., 1994: An order-of-magnitude estimate of the current uplift-rates in Switzerland caused by the Würm alpine deglaciation 545–557. *Eclogae Geologicae Helvetiae*, **87**.
- Gudmundsson, L. and S. I. Seneviratne, 2015: Towards observation-based gridded runoff estimates for Europe. *Hydrology and Earth System Sciences*, **19** (6), 2859–2879.
- Günther, A., J. Wienhöfer, and H. Konietzky, 2012: Automated mapping of rock slope geometry, kinematics and stability with RSS-GIS. *Natural Hazards*, **61** (1), 29–49.
- Gunzburger, Y., V. Merrien-Soukatchoff, and Y. Guglielmi, 2005: Influence of daily surface temperature fluctuations on rock slope stability: case study of the Rochers de Valabres slope (France). *International Journal of Rock Mechanics and Mining Sciences*, **42** (3), 331–349.
- Guyon, I. and A. Elisseeff, 2003: An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research*, **3**, 1157–1182.
- Guyon, I., J. Weston, S. Barnhill, and V. Vapnik, 2002: Gene Selection for Cancer Classification using Support Vector Machines. *Machine Learning*, **46** (1), 389–422.
- Guzzetti, F., A. Carrara, M. Cardinali, and P. Reichenbach, 1999: Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. *Geomorphology*, **31** (1–4), 181–216.
- Gärtner-Roer, I., I. Heinrich, and H. Gärtner, 2013: Wood anatomical analysis of Swiss willow (*Salix helvetica*) shrubs growing on creeping mountain permafrost. *Dendrochronologia*, **31** (2), 97–104.
- Gómez, C., J. C. White, and M. A. Wulder, 2016: Optical remotely sensed time series data for land cover classification: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, **116**, 55–72.
- Hales, T. and J. Roering, 2009: A frost “buzzsaw” mechanism for erosion of the eastern Southern Alps, New Zealand. *Geomorphology*, **107** (3–4), 241–253.
- Hales, T. and J. J. Roering, 2005: Climate-controlled variations in scree production, Southern Alps, New Zealand. *Geology*, **33** (9), 701–704.
- Hales, T. C. and J. J. Roering, 2007: Climatic controls on frost cracking and implications for the evolution of bedrock landscapes. *Journal of Geophysical Research: Earth Surface*, **112** (F2).
- Hall, K., 2006: Perceptions of rock weathering in cold regions: a discussion on space and time attributes of scale. *Géomorphologie*, **3**, 187–196.
- Hall, K. and M.-F. André, 2001: New insights into rock weathering from high-frequency rock temperature data: an Antarctic study of weathering by thermal stress. *Geomorphology*, **41** (1), 23–35.

- Hall, K., J. M. Arocena, J. Boelhouwers, and Z. Liping, 2005: The influence of aspect on the biological weathering of granites: observations from the Kunlun Mountains, China. *Geomorphology*, **67** (1–2), 171–188.
- Hall, K. and C. Thorn, 2011: The historical legacy of spatial scales in freeze–thaw weathering: Misrepresentation and resulting misdirection. *Geomorphology*, **130** (1–2), 83–90.
- Hall, K., C. Thorn, and P. Sumner, 2012: On the persistence of ‘weathering’. *Geomorphology*, **149–150**, 1–10.
- Hall, K. and C. E. Thorn, 2014: Thermal fatigue and thermal shock in bedrock: An attempt to unravel the geomorphic processes and products. *Geomorphology*, **206**, 1–13.
- Hallet, B., J. S. Walder, and C. W. Stubbs, 1991: Weathering by segregation ice growth in microcracks at sustained subzero temperatures: Verification from an experimental study using acoustic emissions. *Permafrost and Periglacial Processes*, **2** (4), 283–300.
- Hamann, U., *et al.*, 2014: Remote sensing of cloud top pressure/height from SEVIRI: analysis of ten current retrieval algorithms. *Atmospheric Measurement Techniques*, **7** (9), 2839–2867.
- Hamdi, S., F. Moyano, S. Sall, M. Bernoux, and T. Chevallier, 2013: Synthesis analysis of the temperature sensitivity of soil respiration from laboratory studies in relation to incubation methods and soil conditions. *Soil Biology and Biochemistry*, **58**, 115–126.
- Hamel, L., 2009: *Model Assessment with ROC Curves*, chap. 204, 1316–1323. 2d ed., Information Science Reference (an imprint of IGI Global), Hershey.
- Haralick, R., K. Shanmugam, and I. Dinstein, 1973: Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics*, **3**, 610–621.
- Harris, R. B., 2010: Rangeland degradation on the Qinghai-Tibetan plateau: A review of the evidence of its magnitude and causes. *Journal of Arid Environments*, **74** (1), 1–12.
- Harrison, S., 2001: On reductionism and emergence in geomorphology. *Transactions of the Institute of British Geographers*, **26** (3), 327–339.
- Hasler, A., S. Gruber, M. Font, and A. Dubois, 2011a: Advective Heat Transport in Frozen Rock Clefts: Conceptual Model, Laboratory Experiments and Numerical Simulation. *Permafrost and Periglacial Processes*, **22** (4), 378–389.
- Hasler, A., S. Gruber, and W. Haeberli, 2011b: Temperature variability and offset in steep alpine rock and ice faces. *The Cryosphere*, **5** (4), 977–988.
- Heckmann, T., L. Hilger, L. Vehling, and M. Becht, 2016: Integrating field measurements, a geomorphological map and stochastic modelling to estimate the spatially distributed rockfall sediment budget of the Upper Kaunertal, Austrian Central Alps. *Geomorphology*, **260**, 16–31.
- Heim, A., 1932: *Bergsturz und Menschenleben*. Fretz & Wasmuth Verlag, Zürich.

- Hencher, S. R., S. G. Lee, T. G. Carter, and L. R. Richards, 2011: Sheeting Joints: Characterisation, Shear Strength and Engineering. *Rock Mechanics and Rock Engineering*, **44** (1), 1–22.
- Henderson, B., E. Bui, C. Moran, and D. Simon, 2005: Australia-wide predictions of soil properties using decision trees. *Geoderma*, **124** (3–4), 383–398.
- Hengl, T., 2009: *Practical Guide to Geostatistical Mapping*. Lulu.com, Amsterdam, Netherlands.
- Hengl, T., J. M. de Jesus, R. A. MacMillan, N. H. Batjes, G. B. M. Heuvelink, E. Ribeiro, A. Samuel-Rosa, B. Kempen, J. G. B. Leenaars, M. G. Walsh, and M. R. Gonzalez, 2014a: SoilGrids1km — Global Soil Information Based on Automated Mapping. *PLOS ONE*, **9** (8), 1–17.
- Hengl, T., G. B. Heuvelink, and D. G. Rossiter, 2007: About regression-kriging: From equations to case studies. *Computers & Geosciences*, **33** (10), 1301–1315.
- Hengl, T., G. B. M. Heuvelink, M. Perčec Tadić, and E. J. Pebesma, 2011: Spatio-temporal prediction of daily temperatures using time-series of MODIS LST images. *Theoretical and Applied Climatology*, **107** (1), 265–277.
- Hengl, T., G. B. M. Heuvelink, M. Perčec Tadić, and E. J. Pebesma, 2012: Spatio-temporal prediction of daily temperatures using time-series of MODIS LST images. *Theoretical and Applied Climatology*, **107** (1), 265–277.
- Hengl, T., B. Kempen, G. Heuvelink, B. Malone, and R. Hannes, 2014b: *GSIF: Global Soil Information Facilities*. URL <http://CRAN.R-project.org/package=GSIF>, R package version 0.4-1.
- Hengl, T., P. Roudier, D. Beaudette, and E. Pebesma, 2015: plotKML: Scientific Visualization of Spatio-Temporal Data. *Journal of Statistical Software*, **63** (1), 1–25.
- Herbst, M., W. Tappe, S. Kummer, and H. Vereecken, 2016: The impact of sieving on heterotrophic respiration response to water content in loamy and sandy topsoils. *Geoderma*, **272**, 73–82.
- Heung, B., H. C. Ho, J. Zhang, A. Knudby, C. E. Bulmer, and M. G. Schmidt, 2016: An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. *Geoderma*, **265**, 62–77.
- Heuvelink, G., D. Griffith, T. Hengl, and S. Melles, 2012: Sampling Design Optimization for Space-Time Kriging. *Spatio-Temporal Design*, W. Jorgeteu, Müller, Ed., John Wiley & Sons, Ltd, 207–230.
- Heuvelink, G. and R. Webster, 2001: Modelling soil variation: past, present, and future. *Geoderma*, **100** (3–4), 269–301, developments and Trends in Soil Science.
- Hijmans, R., J. van Etten, M. Mattiuzzi, M. Sumner, J. Greenberg, O. Lamigueiro, A. Bevan, E. Racine, and A. Shortridge, 2014: *raster: Geographic Data Analysis and Modeling*. URL <http://CRAN.R-project.org/package=raster>, R package version 2.3-12.

- Hijmans, R. J., 2015: *raster: Geographic Data Analysis and Modeling*. URL <http://CRAN.R-project.org/package=raster>, R package version 2.4-18.
- Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis, 2005: Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25** (15), 1965–1978.
- Hijmans, R. J., S. Phillips, J. Leathwick, and J. Elith, 2015: *dismo: Species Distribution Modeling*. URL <http://CRAN.R-project.org/package=dismo>, R package version 1.0-12.
- Hilhorst, M. A., 2000: A Pore Water Conductivity Sensor. *Soil Science Society of America Journal*, **64**, 1922–1925.
- Hinchliffe, S. and C. K. Ballantyne, 1999: Talus accumulation and Rockwall retreat, Trotternish, isle of Skye, Scotland. *Scottish Geographical Journal*, **115** (1), 53–70.
- Ho, H. C., A. Knudby, P. Sirovyak, Y. Xu, M. Hodul, and S. B. Henderson, 2014: Mapping maximum urban air temperature on hot summer days. *Remote Sensing of Environment*, **154**, 38–45.
- Hoek, E. and E. Brown, 1997: Practical estimates of rock mass strength. *International Journal of Rock Mechanics and Mining Sciences*, **34** (8), 1165–1186.
- Hoffmann, T., T. Müller, E. A. Johnson, and Y. E. Martin, 2013: Postglacial adjustment of steep, low-order drainage basins, Canadian Rocky Mountains. *Journal of Geophysical Research: Earth Surface*, **118** (4), 2568–2584.
- Hofstra, N., M. Haylock, M. New, P. Jones, and C. Frei, 2008: Comparison of six methods for the interpolation of daily, European climate data. *Journal of Geophysical Research: Atmospheres*, **113** (D21), d21110.
- Hong, Y., K.-L. Hsu, S. Sorooshian, and X. Gao, 2004: Precipitation Estimation from Remotely Sensed Imagery Using an Artificial Neural Network Cloud Classification System. *Journal of Applied Meteorology*, **43** (12), 1834–1853.
- Hosmer, D. and S. Lemeshow, (Eds.) , 2005: *Applied Logistic Regression*. John Wiley & Sons, New York.
- Hothorn, T., 2017: CRAN Task View: Machine Learning & Statistical Learning. URL <https://CRAN.R-project.org/view=MachineLearning>.
- Hou, A. Y., R. K. Kakar, S. Neeck, A. Azarbarzin, C. D. Kummerow, M. Kojima, R. Oki, K. Nakamura, and T. Iguchi, 2014: The Global Precipitation Measurement Mission. *Bulletin of the American Meteorological Society*, **95**, 701–722.
- Hou, X., 2001: *Vegetation Atlas of China (1: 1,000,000)*. Beijing, China: Science Press.
- Hsieh, W. W., 2009: *Machine Learning Methods in the Environmental Sciences: Neural Networks and Kernels*. University of British Columbia, Vancouver.
- Hsu, K., X. Gao, S. Sorooshian, and H. Gupta, 1997: Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks. *Journal of Applied Meteorology*, **36** (9), 1176–1190.

- Hu, Q., W. Wu, T. Xia, Q. Yu, P. Yang, Z. Li, and Q. Song, 2013: Exploring the Use of Google Earth Imagery and Object-Based Methods in Land Use/Cover Mapping. *Remote Sensing*, **5** (11), 6026–6042.
- Huang, C., L. S. Davis, and J. R. G. Townshend, 2002: An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*, **23** (4), 725–749.
- Huang, R., C. Zhang, J. Huang, D. Zhu, L. Wang, and J. Liu, 2015: Mapping of Daily Mean Air Temperature in Agricultural Regions Using Daytime and Nighttime Land Surface Temperatures Derived from TERRA and AQUA MODIS Data. *Remote Sensing*, **7** (7), 8728.
- Huang, Z., B. J. Turner, S. J. Dury, I. R. Wallis, and W. J. Foley, 2004: Estimating foliage nitrogen concentration from HYMAP data using continuum removal analysis. *Remote Sensing of Environment*, **93** (1–2), 18–29.
- Hudak, A. T. and C. A. Wessman, 1998: Textural Analysis of Historical Aerial Photography to Characterize Woody Plant Encroachment in South African Savanna. *Remote Sensing of Environment*, **66** (3), 317–330.
- Hudak, A. T. and C. A. Wessman, 2001: Textural analysis of high resolution imagery to quantify bush encroachment in Madikwe Game Reserve, South Africa, 1955–1996. *International Journal of Remote Sensing*, **22**, 2731–2740.
- Huffman, G., D. Bolvin, D. Braithwaite, K. Hsu, R. Joyce, and P. Xie, 2014: GPM L3 IMERG Late Half Hourly 0.1 degree x 0.1 degree Precipitation V03, Greenbelt, MD, Goddard Earth Sciences Data and Information Services Center (GES DISC). URL ftp://gpm1.gesdisc.eosdis.nasa.gov/data/s4pa/GPM_L3/GPM_3IMERGHH.03/, (last access: 15 June, 2015), doi: 10.5067/GPM/IMERG/HH/3B.
- Huggel, C., J. J. Clague, and O. Korup, 2012: Is climate change responsible for changing landslide activity in high mountains? *Earth Surface Processes and Landforms*, **37** (1), 77–91.
- Huisman, J. A., J. J. J. C. Snepvangers, W. Bouten, and G. B. M. Heuvelink, 2003: Monitoring Temporal Development of Spatial Soil Water Content Variation. *Vadose Zone Journal*, **2**, 519–529.
- Hungr, O., S. G. Evans, and J. Hazzard, 1999: Magnitude and frequency of rock falls and rock slides along the main transportation corridors of southwestern British Columbia. *Canadian Geotechnical Journal*, **36** (2), 224–238.
- Hébrard, O., M. Voltz, P. Andrieux, and R. Moussa, 2006: Spatio-temporal distribution of soil surface moisture in a heterogeneously farmed Mediterranean catchment. *Journal of Hydrology*, **329** (1–2), 110–121.
- Im, J., S. Park, J. Rhee, J. Baik, and M. Choi, 2016: Downscaling of AMSR-E soil moisture with MODIS products using machine learning approaches. *Environmental Earth Sciences*, **75** (15), 1120.
- IPCC, 2014: *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. IPCC, Geneva, Switzerland.

- IPWG, 2016: IPWG South African Validation. URL http://rsmc.weathersa.co.za/IPWG/ipwgsa_qlooks.html, (last access: 1 February 2017).
- Islam, T., M. A. Rico-Ramirez, P. K. Srivastava, and Q. Dai, 2014: Non-parametric rain/no rain screening method for satellite-borne passive microwave radiometers at 19–85 GHz channels with the Random Forests algorithm. *International Journal of Remote Sensing*, **35** (9), 3254–3267.
- ISO 10694, 1995: *Soil quality–determination of organic and total carbon after dry combustion (elemental analysis)*. ISO-Secretariat, Geneva, Switzerland.
- Itano, S. and H. Tomimatsu, 2011: Reflectance spectra for monitoring green herbage mass in Zoysia-dominated pastures. *Grassland Science*, **57** (1), 9–17.
- Ivy-Ochs, S., H. Kerschner, M. Maisch, M. Christl, P. W. Kubik, and C. Schlüchter, 2009: Latest Pleistocene and Holocene glacier variations in the European Alps. *Quaternary Science Reviews*, **28** (21–22), 2137–2149.
- Jaboyedoff, M., R. Metzger, T. Oppikofer, R. Couture, M.-H. Derron, J. Locat, and D. Turmel, 2007: New insight techniques to analyze rock-slope relief using DEM and 3D-imaging cloud points. *Rock Mechanics: Meeting Society's Challenges and Demands, Two Volume Set*, CRC Press, 61–68.
- James, G., D. Witten, T. Hastie, and R. Tibshirani, 2013: *An Introduction to Statistical Learning: with Applications in R*. 1st ed., Springer, New York.
- Janatian, N., M. Sadeghi, S. H. Sanaeinejad, E. Bakhshian, A. Farid, S. M. Hashemini, and S. Ghazanfari, 2017: A statistical framework for estimating air temperature using MODIS land surface temperature data. *International Journal of Climatology*, **37** (3), 1181–1194.
- Janik, L. and J. Skjemstad, 1995: Characterization and analysis of soils using mid-infrared partial least-squares .2. Correlations with some laboratory data. *Soil Res.*, **33** (4), 637–650.
- Janke, J. R., 2005: Modeling past and future alpine permafrost distribution in the Colorado Front Range. *Earth Surface Processes and Landforms*, **30** (12), 1495–1508.
- Jarvis, C. H. and N. Stuart, 2001: A Comparison among Strategies for Interpolating Maximum and Minimum Daily Air Temperatures. Part II: The Interaction between Number of Guiding Variables and the Type of Interpolation Method. *Journal of Applied Meteorology*, **40** (6), 1075–1084.
- Jin, M. and R. E. Dickinson, 1999: Interpolation of surface radiative temperature measured from polar orbiting satellites to a diurnal cycle: 1. Without clouds. *Journal of Geophysical Research: Atmospheres*, **104** (D2), 2105–2116.
- Jin, Z. N., Q. L. Zhuang, J. S. He, T. X. Luo, and Y. Shi, 2013: Phenology shift from 1989 to 2008 on the Tibetan Plateau: an analysis with a process-based soil physical model and remote sensing data. *Climatic Change*, **119** (2), 435–449.
- Jing, W., Y. Yang, X. Yue, and X. Zhao, 2016: A Comparison of Different Regression Algorithms for Downscaling Monthly Satellite-Based Precipitation over North China. *Remote Sensing*, **8** (10), 835.

- Joseph, L. and L. Devadas, 2015: Detection of rooftop regions in rural areas using Support Vector Machine. *International Journal of Scientific Research Engineering & Technology*, **4** (5), 549–553.
- Jost, G., G. Heuvelink, and A. Papritz, 2005: Analysing the space–time distribution of soil water storage of a forest ecosystem using spatio-temporal kriging. *Geoderma*, **128** (3–4), 258–273.
- Jung, T., E. Ruprecht, and F. Wagner, 1998: Determination of Cloud Liquid Water Path over the Oceans from Special Sensor Microwave/Imager (SSM/I) Data Using Neural Networks. *Journal of Applied Meteorology*, **37** (8), 832–844.
- Jäckli, H., 1957: *Gegenwartsgeologie des bündnerischen Rheingebietes: Ein Beitrag zur exogenen Dynamik alpiner Gebirgslandschaften (Exogene dynamics of an Alpine landscape)*, Beiträge zur Geologie der Schweiz, Geotechnische Serie, Vol. 36. Kümmerly & Frey AG, Bern.
- Kaiser, H. F., 1958: The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, **23** (3), 187–200.
- Kanevski, M., A. Pozdnukhov, and V. Timonin, 2009: *Machine Learning for Spatial Environmental Data: Theory, Applications and Software*. 1st ed., EPFL Press.
- Kaptué, A. T., N. P. Hanan, L. Prihodko, and J. A. Ramirez, 2015: Spatial and temporal characteristics of rainfall in Africa: Summary statistics for temporal downscaling. *Water Resources Research*, **51** (4), 2668–2679.
- Karatzoglou, A., T. U. Wien, A. Smola, K. Hornik, and W. Wien, 2004: kernlab – An S4 Package for Kernel Methods in R. *Journal of Statistical Software*, **11**, 1–20.
- Kavzoglu, T. and I. Colkesen, 2009: A kernel functions analysis for support vector machines for land cover classification. *International Journal of Applied Earth Observation and Geoinformation*, **11** (5), 352–359.
- Ke, Y., J. Im, S. Park, and H. Gong, 2016: Downscaling of MODIS One Kilometer Evapotranspiration Using Landsat-8 Data and Machine Learning Approaches. *Remote Sensing*, **8** (3), 215.
- Keith, H., K. Jacobsen, and R. Raison, 1997: Effects of soil phosphorus availability, temperature and moisture on soil respiration in Eucalyptus pauciflora forest. *Plant and Soil*, **190** (1), 127–141.
- Kelly, M. A., P. W. Kubik, F. Von Blanckenburg, and C. Schlüchter, 2004: Surface exposure dating of the Great Aletsch Glacier Egesen moraine system, western Swiss Alps, using the cosmogenic nuclide ^{10}Be . *Journal of Quaternary Science*, **19** (5), 431–441.
- Kerschen, G. and J. Golival, 2004: Feature extraction using auto-associative neural networks. *Smart Materials and Structures*, **13** (1), 211.
- Kidd, C., P. Bauer, J. Turk, G. J. Huffman, R. Joyce, K.-L. Hsu, and D. Braithwaite, 2011: Intercomparison of High-Resolution Precipitation Products over Northwest Europe. *Journal of Hydrometeorology*, **13** (1), 67–83.

- Kidd, C. and G. Huffman, 2011: Global precipitation measurement. *Meteorological Applications*, **18** (3), 334–353.
- Kidd, C. and V. Levizzani, 2011: Status of satellite precipitation retrievals. *Hydrology and Earth System Sciences*, **15** (4), 1109–1116.
- Kilibarda, M., T. Hengl, G. B. M. Heuvelink, B. Gräler, E. Pebesma, M. Perčec Tadić, and B. Bajat, 2014: Spatio-temporal interpolation of daily temperatures for global land areas at 1 km resolution. *Journal of Geophysical Research: Atmospheres*, **119** (5), 2294–2313.
- Kirschbaum, M. U., 1995: The temperature dependence of soil organic matter decomposition, and the effect of global warming on soil organic C storage. *Soil Biology and Biochemistry*, **27** (6), 753–760.
- Kirschbaum, M. U. F., 2006: The temperature dependence of organic-matter decomposition—still a topic of debate. *Soil Biology and Biochemistry*, **38** (9), 2510–2518.
- Kleinbaum, D., L. Kupper, K. Muller, and A. Nizam, 1998: *Applied Regression Analysis and Other Multivariable Methods*. Duxbury Press, Pacific Grove.
- Kniffka, A., M. Stengel, and R. Hollmann, 2014: *SEVIRI cloud mask dataset - Edition 1 - 15 minutes resolution*. Satellite Application Facility on Climate Monitoring. EUMETSAT Satellite Application Facility on Climate Monitoring (CM SAF).
- Knox, N., S. Grunwald, M. McDowell, G. Bruland, D. Myers, and W. Harris, 2015: Modelling soil carbon fractions with visible near-infrared (VNIR) and mid-infrared (MIR) spectroscopy. *Geoderma*, **239**, 229–239.
- Koch, O., D. Tscherko, and E. Kandeler, 2007: Temperature sensitivity of microbial respiration, nitrogen mineralization, and potential soil enzyme activities in organic alpine soils. *Global Biogeochemical Cycles*, **21** (4).
- Kokaly, R. F. and R. N. Clark, 1999: Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression. *Remote Sensing of Environment*, **67** (3), 267–287.
- Kotchenova, S. Y., E. F. Vermote, R. Matarrese, and J. Frank J. Klemm, 2006: Validation of a vector version of the 6S radiative transfer code for atmospheric correction of satellite data. Part I: Path radiance. *Applied Optics*, **45** (26), 6762–6774.
- Krautblatter, M. and R. Dikau, 2007: Towards a uniform Concept for the Comparison and Extrapolation of Rockwall Retreat and Rockfall Supply. *Geografiska Annaler: Series A, Physical Geography*, **89** (1), 21–40.
- Krautblatter, M., D. Funk, and F. K. Günzel, 2013: Why permafrost rocks become unstable: a rock–ice–mechanical model in time and space. *Earth Surface Processes and Landforms*, **38** (8), 876–887.
- Krautblatter, M. and C. Hauck, 2007: Electrical resistivity tomography monitoring of permafrost in solid rock walls. *Journal of Geophysical Research: Earth Surface*, **112** (F2).

- Krautblatter, M. and J. R. Moore, 2014: Rock slope instability and erosion: toward improved process understanding. *Earth Surface Processes and Landforms*, **39** (9), 1273–1278.
- Krautblatter, M., M. Moser, L. Schrott, J. Wolf, and D. Morche, 2012: Significance of rockfall magnitude and carbonate dissolution for rock slope erosion and geomorphic work on Alpine limestone cliffs (Reintal, German Alps). *Geomorphology*, **167–168**, 21–34.
- Krinner, G., N. Viovy, N. de Noblet-Ducoudré, J. Ogée, J. Polcher, P. Friedlingstein, P. Ciais, S. Sitch, and I. C. Prentice, 2005: A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system. *Global Biogeochemical Cycles*, **19** (1).
- Krogh, A. and J. A. Hertz, 1992: A Simple Weight Decay Can Improve Generalization. *Advances in Neural Information Processing Systems 4*, Morgan Kaufmann, 950–957.
- Kruger, A. C., (Ed.), 2007: *Climate of South Africa, Precipitation*, Vol. Report No. WS47, South African Weather Service, Pretoria, South Africa.
- Kuhn, M., 2014a: *caret: Classification and Regression Training*. URL <http://CRAN.R-project.org/package=caret>, R package version 6.0-29.
- Kuhn, M., 2014b: Optimizing probability thresholds for class imbalances. URL <http://www.r-bloggers.com/optimizing-probability-thresholds-for-class-imbalances/>.
- Kuhn, M., 2016a: *caret: Classification and Regression Training*. URL <https://CRAN.R-project.org/package=caret>, R package version 6.0-71.
- Kuhn, M., 2016b: *caret: Classification and Regression Training*. URL <https://CRAN.R-project.org/package=caret>, R package version 6.0-68.
- Kuhn, M., 2017: *caret: Classification and Regression Training*. URL <https://CRAN.R-project.org/package=caret>, R package version 6.0-76.
- Kuhn, M. and K. Johnson, 2013a: *Applied Predictive Modeling*. 1st ed., Springer, New York.
- Kuhn, M. and K. Johnson, 2013b: *Applied Predictive Modeling*, chap. 7.1 Neural Networks, 141 – 145. 1st ed., Springer, New York.
- Kuhn, M., S. Weston, C. Keefer, and N. C. C. code for Cubist by Ross Quinlan, 2014: *Cubist: Rule- and Instance-Based Regression Modeling*. URL <https://CRAN.R-project.org/package=Cubist>, R package version 0.0.18.
- Kühnlein, M., 2014: A machine learning based 24-h-technique for an area-wide rainfall retrieval using MSG SEVIRI data over Central Europe. Ph.D. thesis, Fachbereich Geographie der Philipps-Universität Marburg.
- Kühnlein, M., T. Appelhans, B. Thies, and T. Nauss, 2014a: Improving the accuracy of rainfall rates from optical satellite sensors with machine learning - A random forests-based approach applied to MSG SEVIRI. *Remote Sensing of Environment*, **141**, 129–143.

- Kühnlein, M., T. Appelhans, B. Thies, and T. Nauss, 2014b: Precipitation Estimates from MSG SEVIRI Daytime, Nighttime, and Twilight Data with Random Forests. *Journal of Applied Meteorology and Climatology*, **53** (11), 2457–2480.
- Kühnlein, M., B. Thies, T. Nauss, and J. Bendix, 2010: Rainfall-Rate Assignment Using MSG SEVIRI Data - A Promising Approach to Spaceborne Rainfall-Rate Retrieval for Midlatitudes. *Journal of Applied Meteorology and Climatology*, **49** (7), 1477–1495.
- Kyriakidis, P. C. and A. G. Journel, 1999: Geostatistical Space-Time Models: A Review. *Mathematical Geology*, **31** (6), 651–684.
- Labhart, T., 2009: *Geologie der Schweiz*. Ott Verlag, Thun.
- Lakshmi, V., 2013: Remote Sensing of Soil Moisture. *ISRN Soil Science*, **2013**, 1–33.
- Land Processes Distributed Active Archive Center (LP DAAC), 2013: MODIS Level 3 Land Surface Temperature and Emissivity. Version 3. URL <https://lpdaac.usgs.gov>, (last access: 26 April 2016).
- Langella, G., A. Basile, A. Bonfante, and F. Terribile, 2010: High-resolution space-time rainfall analysis using integrated ANN inference systems. *Journal of Hydrology*, **387** (3–4), 328–342.
- Lary, D., F. Faruque, N. Malakar, A. Moore, B. Roscoe, Z. Adams, and Y. Eggleston, 2014: Estimating the global abundance of ground level presence of particulate matter (PM_{2.5}). *Geospatial Health*, **8** (3), 611–630.
- Lary, D. J., A. H. Alavi, A. H. Gandomi, and A. L. Walker, 2016: Machine learning in geosciences and remote sensing. *Geoscience Frontiers*, **7** (1), 3–10.
- Lawrence, R. L., S. D. Wood, and R. L. Sheley, 2006: Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (randomForest). *Remote Sensing of Environment*, **100** (3), 356–362.
- Lazri, M., S. Ameur, and Y. Mohia, 2014: Instantaneous rainfall estimation using neural network from multispectral observations of SEVIRI radiometer and its application in estimation of daily and monthly rainfall. *Advances in Space Research*, **53** (1), 138–155.
- Lazzara, M. A., G. A. Weidner, L. M. Keller, J. E. Thom, and J. J. Cassano, 2012: Antarctic Automatic Weather Station Program: 30 Years of Polar Observation. *Bulletin of the American Meteorological Society*, **93** (10), 1519–1537.
- Lee, S. and T. Sambath, 2006: Landslide susceptibility mapping in the Damrei Romel area, Cambodia using frequency ratio and logistic regression models. *Environmental Geology*, **50** (6), 847–855.
- Lefèvre, R., P. Barré, F. Moyano, B. Christensen, G. Bardoux, T. Eglin, C. Girardin, S. Houot, T. Kätterer, F. van Oort, and C. Chenu, 2014: Higher temperature sensitivity for stable than for labile soil organic carbon - Evidence from incubations of long-term bare fallow soils. *Global Change Biology*, **20**, 633–640.

- Legendre, P., 1993: Spatial Autocorrelation: Trouble or New Paradigm? *Ecology*, **74** (6), 1659–1673.
- Lehnert, L., H. Meyer, N. Meyer, C. Reudenbach, and J. Bendix, 2014: A hyperspectral indicator system for rangeland degradation on the Tibetan Plateau: A case study towards spaceborne monitoring. *Ecological Indicators*, **39**, 54–64.
- Lehnert, L. W., H. Meyer, and J. Bendix, 2015a: *hsdar: Manage, analyse and simulate hyperspectral data in R*. R package version 0.3.0.
- Lehnert, L. W., H. Meyer, N. Meyer, C. Reudenbach, and J. Bendix, 2013: Assessing pasture quality and degradation status using hyperspectral imaging: a case study from western Tibet. *Proc. SPIE 8887, Remote Sensing for Agriculture, Ecosystems, and Hydrology XV, 88870I (16 October 2013)*.
- Lehnert, L. W., H. Meyer, Y. Wang, G. Mieke, B. Thies, C. Reudenbach, and J. Bendix, 2015b: Retrieval of grassland plant coverage on the Tibetan Plateau based on a multi-scale, multi-sensor and multi-method approach. *Remote Sensing of Environment*, **164**, 197–207.
- Lehnert, L. W., K. Wesche, K. Trachte, C. Reudenbach, and J. Bendix, 2016: Climate variability rather than overstocking causes recent large scale cover changes of Tibetan pastures. *Scientific Reports*, **6**, 24367.
- Leith, K., J. R. Moore, F. Amann, and S. Loew, 2014: Subglacial extensional fracture development and implications for Alpine Valley evolution. *Journal of Geophysical Research: Earth Surface*, **119** (1), 62–81.
- Lensky, I. M. and D. Rosenfeld, 1997: Estimation of Precipitation Area and Rain Intensity Based on the Microphysical Properties Retrieved from NOAA AVHRR Data. *Journal of Applied Meteorology*, **36** (3), 234–242.
- Levizzani, V., R. Amorati, and F. Meneguzzo, 2002: A Review of Satellite-based Rainfall Estimation Methods. Tech. rep., European Commission Project MUSIC Report (EVK1-CT-2000-00058).
- Levizzani, V., J. Schmetz, H. J. Lutz, J. Kerkmann, P. P. Alberoni, and M. Cervino, 2001: Precipitation estimations from geostationary orbit and prospects for METEOSAT Second Generation. *Meteorological Applications*, **8** (1), 23–41.
- Lewis, W. V., 1954: Pressure Release and Glacial Erosion. *Journal of Glaciology*, **2** (16), 417–422.
- Li, J., A. D. Heap, A. Potter, and J. J. Daniell, 2011: Application of machine learning methods to spatial interpolation of environmental variables. *Environmental Modelling & Software*, **26** (12), 1647–1659.
- Liaw, A. and M. Wiener, 2002: Classification and Regression by randomForest. *R News*, **2** (3), 18–22.
- Lichtenthaler, H. and A. Wellburn, 1983: Determinations of total carotenoids and chlorophylls a and b of leaf extracts in different solvents. *Biochemical Society Transactions*, **11** (5), 591–592.

- Ließ, M., J. Schmidt, and B. Glaser, 2016: Improving the Spatial Prediction of Soil Organic Carbon Stocks in a Complex Tropical Mountain Landscape by Methodological Specifications in Machine Learning Approaches. *PLOS ONE*, **11** (4), 1–22.
- Lin, S., N. J. Moore, J. P. Messina, M. H. DeVisser, and J. Wu, 2012: Evaluation of estimating daily maximum and minimum air temperature with MODIS data in east Africa. *International Journal of Applied Earth Observation and Geoinformation*, **18**, 128–140.
- Linker, R., M. Weiner, I. Shmulevich, and A. Shaviv, 2006: Nitrate Determination in Soil Pastes using Attenuated Total Reflectance Mid-infrared Spectroscopy: Improved Accuracy via Soil Identification. *Biosystems Engineering*, **94** (1), 111–118.
- Liu, H., K. C. Jezek, B. Li, and Z. Zhao., 2015: Radarsat Antarctic Mapping Project Digital Elevation Model, Version 2.
- Liu, Y., N. V. Chawla, M. P. Harper, E. Shriberg, and A. Stolcke, 2006: A study in machine learning from imbalanced data for sentence boundary detection in speech. *Computer Speech & Language*, **20** (4), 468–494.
- Liu, Y., D.-G. Xi, Z.-L. Li, and C.-X. Shi, 2014: Analysis and Application of the Relationship between Cumulonimbus (Cb) Cloud Features and Precipitation Based on FY-2C Image. *Atmosphere*, **5** (2), 211–229.
- Liu, Y., Y. Zha, J. Gao, and S. Ni, 2004: Assessment of grassland degradation near Lake Qinghai, West China, using Landsat TM and in situ reflectance spectra data. *International Journal of Remote Sensing*, **25** (20), 4177–4189.
- Long, R. J., S. O. Apori, F. B. Castro, and E. R. Orskov, 1999: Feed value of native forages of the Tibetan Plateau of China. *Animal Feed Science and Technology*, **80** (2), 101–113.
- Lopresti, M. F., C. M. D. Bella, and A. J. Degioanni, 2015: Relationship between MODIS-NDVI data and wheat yield: A case study in Northern Buenos Aires province, Argentina. *Information Processing in Agriculture*, **2** (2), 73–84.
- Loye, A., M. Jaboyedoff, and A. Pedrazzini, 2009: Identification of potential rockfall source areas at a regional scale using a DEM-based geomorphometric analysis. *Natural Hazards and Earth System Sciences*, **9** (5), 1643–1653.
- Loye, A., A. Pedrazzini, J. I. Theule, M. Jaboyedoff, F. Liébault, and R. Metzger, 2012: Influence of bedrock structures on the spatial pattern of erosional landforms in small alpine catchments. *Earth Surface Processes and Landforms*, **37** (13), 1407–1423.
- Luan, J., S. Liu, J. Wang, and X. Zhu, 2013: Factors affecting spatial variation of annual apparent Q₁₀ of soil respiration in two warm temperate forests. *PLoS ONE*, **8**, e64167.
- Ludwig, A., H. Meyer, and T. Nauss, 2016: Automatic classification of Google Earth images for a larger scale monitoring of bush encroachment in South Africa. *International Journal of Applied Earth Observation and Geoinformation*, **50**, 89–94.

- Ludwig, B., P. Khanna, J. Bauhus, and P. Hopmans, 2002: Near infrared spectroscopy of forest soils to determine chemical and biological properties related to soil sustainability. *Forest Ecology and Management*, **171** (1), 121–132.
- Ludwig, B., R. Nitschke, T. Terhoeven-Urselmans, K. Michel, and H. Flessa, 2008: Use of mid-infrared spectroscopy in the diffuse-reflectance mode for the prediction of the composition of organic matter in soil and litter. *Journal of Plant Nutrition and Soil Science*, **171** (3), 384–391.
- Ly, S., C. Charles, and A. Degré, 2011: Geostatistical interpolation of daily rainfall at catchment scale: the use of several variogram models in the Ourthe and Ambleve catchments, Belgium. *Hydrology and Earth System Sciences*, **15** (7), 2259–2274.
- Magnin, F., M. Krautblatter, P. Deline, L. Ravanel, E. Malet, and A. Bevington, 2015: Determination of warm, sensitive permafrost areas in near-vertical rockwalls and evaluation of distributed models by electrical resistivity tomography. *Journal of Geophysical Research: Earth Surface*, **120** (5), 745–762.
- Malone, B., A. McBratney, B. Minasny, and G. Laslett, 2009: Mapping continuous depth functions of soil carbon storage and available water capacity. *Geoderma*, **154** (1–2), 138–152.
- Manhique, A. J., C. J. C. Reason, B. Silinto, J. Zucula, I. Raiva, F. Congolo, and A. F. Mavume, 2015: Extreme rainfall and floods in southern Africa in January 2013 and associated circulation patterns. *Natural Hazards*, **77** (2), 679–691.
- Margenot, A. J., F. J. Calderón, T. M. Bowles, S. J. Parikh, and L. E. Jackson, 2015: Soil Organic Matter Functional Group Composition in Relation to Organic Carbon, Nitrogen, and Phosphorus Fractions in Organically Managed Tomato Fields. *Soil Science Society of America Journal*, **79**, 772–782.
- Marquínez, J., R. Menéndez Duarte, P. Farias, and M. Jiménez Sánchez, 2003: Predictive GIS-Based Model of Rockfall Activity in Mountain Cliffs. *Natural Hazards*, **30** (3), 341–360.
- Martz, L. and J. Garbrecht, 1992: Numerical definition of drainage network and subcatchment areas from Digital Elevation Models. *Computers & Geosciences*, **18** (6), 747–761.
- Matasci, B., M. Jaboyedoff, A. Loe, A. Pedrazzini, M.-H. Derron, and G. Pedrozzi, 2015: Impacts of fracturing patterns on the rockfall susceptibility and erosion rate of stratified limestone. *Geomorphology*, **241**, 83–97.
- Matsuoka, N., 1994: Matsuoka, N., 1994. Diurnal freeze–thaw depth in rockwalls - Field-measurements and theoretical considerations. *Earth Surface Processes and Landforms*, **19**, 423–435.
- Matsuoka, N., 2001: Microgelivation versus macrogelivation: towards bridging the gap between laboratory and field frost weathering. *Permafrost and Periglacial Processes*, **12** (3), 299–313.
- Matsuoka, N. and J. Murton, 2008: Frost weathering: recent advances and future directions. *Permafrost and Periglacial Processes*, **19** (2), 195–210.

- Matsuoka, N. and H. Sakai, 1999: Rockfall activity from an alpine cliff during thawing periods. *Geomorphology*, **28** (3–4), 309–328.
- Mazzocchi, F., 2015: Could Big Data be the end of theory in science? *EMBO Reports*, **16** (10), 1250–1255.
- McBratney, A., M. M. Santos, and B. Minasny, 2003: On digital soil mapping. *Geoderma*, **117** (1–2), 3–52.
- McColl, S. T., 2012: Paraglacial rock-slope stability. *Geomorphology*, **153–154**, 1–16.
- McKenzie, N. J. and P. J. Ryan, 1999: Spatial prediction of soil properties using environmental correlation. *Geoderma*, **89** (1–2), 67–94.
- Merk, C., J. Cermak, and J. Bendix, 2011: Retrieval of optical and microphysical cloud properties from Meteosat SEVIRI data at night – a feasibility study based on radiative transfer calculations. *Remote Sensing Letters*, **2** (4), 357–366.
- Messenzehl, K. and D. Draebing, 2015: Multidisciplinary investigations on coupled rotalus-systems (Turtmann valley, Swiss Alps). *Geophysical Research Abstracts*, Vol. 17.
- Messenzehl, K., T. Hoffmann, and R. Dikau, 2014: Sediment connectivity in the high-alpine valley of Val Mütsch, Swiss National Park — linking geomorphic field mapping with geomorphometric modelling. *Geomorphology*, **221**, 215–229.
- Messenzehl, K., H. Meyer, J.-C. Otto, T. Hoffmann, and R. Dikau, 2017: Regional-scale controls on the spatial activity of rockfalls (Turtmann Valley, Swiss Alps) – A multivariate modeling approach. *Geomorphology*, **287**, 29–45.
- Meyer, H., J. Drönner, and T. Nauss, 2017a: Satellite-based high-resolution mapping of rainfall over southern Africa. *Atmospheric Measurement Techniques*, **10** (6), 2009–2019.
- Meyer, H., M. Katurji, T. Appelhans, M. U. Müller, T. Nauss, P. Roudier, and P. Zawar-Reza, 2016a: Mapping Daily Air Temperature for Antarctica Based on MODIS LST. *Remote Sensing*, **8** (9), 732.
- Meyer, H., M. Kühnlein, T. Appelhans, and T. Nauss, 2016b: Comparison of four machine learning algorithms for their applicability in satellite-based optical rainfall retrievals. *Atmospheric Research*, **169**, Part B, 424–433.
- Meyer, H., M. Kühnlein, C. Reudenbach, and T. Nauss, 2017b: Revealing the potential of spectral and textural predictor variables in a neural network-based rainfall retrieval technique. *Remote Sensing Letters*, **8** (7), 647–656.
- Meyer, H., L. W. Lehnert, Y. Wang, C. Reudenbach, T. Nauss, and J. Bendix, 2017c: From local spectral measurements to maps of vegetation cover and biomass on the Qinghai-Tibet-Plateau: Do we need hyperspectral information? *International Journal of Applied Earth Observation and Geoinformation*, **55**, 21–31.
- Meyer, N., G. Welp, and W. Amelung, under review: What determines the temperature sensitivity of soil respiration (Q₁₀) at catchment scale? *Global Biogeochemical Cycles*.

- Micheletti, N., L. Foresti, S. Robert, M. Leuenberger, A. Pedrazzini, M. Jaboyedoff, and M. Kanevski, 2014: Machine Learning Feature Selection Methods for Landslide Susceptibility Mapping. *Mathematical Geosciences*, **46** (1), 33–57.
- Michoud, C., M.-H. Derron, P. Horton, M. Jaboyedoff, F.-J. Baillifard, A. Loye, P. Nicolet, A. Pedrazzini, and A. Queyrel, 2012: Rockfall hazard and risk assessments along roads at a regional scale: example in Swiss Alps. *Natural Hazards and Earth System Sciences*, **12** (3), 615–629.
- Miehe, G., 2004: Hochland von Tibet. *Gebirge der Erde: Landschaft, Klima, Pflanzenwelt*, C. A. Burga, F. Klötzli, and G. Grabherr, Eds., Ulmer Verlag.
- Miehe, G., K. Bach, S. Miehe, J. Kluge, Y. Yongping, L. Duo, S. Co, and K. Wesche, 2011a: Alpine steppe plant communities of the Tibetan highlands. *Applied Vegetation Science*, **14** (4), 547–560.
- Miehe, G., K. Kaiser, S. Co, Z. Xinquan, and L. Jianquan, 2008a: Geo-ecological transect studies in northeast Tibet (Qinghai, China) reveal human-made mid-Holocene environmental changes in the upper Yellow River catchment changing forest to grassland. *Erdkunde*, **62** (3), 187–199.
- Miehe, G., S. Miehe, K. Bach, J. Nölling, J. Hanspach, C. Reudenbach, K. Kaiser, K. Wesche, V. Mosbrugger, Y. Yang, and Y. Ma, 2011b: Plant communities of central Tibetan pastures in the alpine steppe/Kobresia pygmaea ecotone. *Journal of Arid Environments*, **75** (8), 711–723.
- Miehe, G., S. Miehe, K. Kaiser, L. Jianquan, and X. Zhao, 2008b: Status and dynamics of the Kobresia pygmaea ecosystem on the Tibetan Plateau. *AMBIO*, **37** (4), 272–279.
- Miehe, G., S. Miehe, K. Kaiser, C. Reudenbach, L. Behrendes, L. Duo, and F. Schluetz, 2009: How old is pastoralism in Tibet? An ecological approach to the making of a Tibetan landscape. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **276** (1–4), 130–147.
- Miller, H. and M. F. Goodchild, 2015: Data-driven geography. *GeoJournal*, **80** (4), 449–461.
- Miller, R., (Ed.) , 1994: *Mapping the Diversity of Nature*. Springer-Science+Business Media, B.V.
- Miska, L. and H. Jan, 2005: Evaluation of current statistical approaches for predictive geomorphological mapping. *Geomorphology*, **67** (3–4), 299–315.
- Mitas, L. and H. Mitasova, 1999: Spatial interpolation. *Geographical Information Systems: Principles, Techniques, Management and Applications*, P. Longley, M. Goodchild, D. Maguire, and D. Rhind, Eds., Wiley, Vol. 1, 481–492.
- Moore, J. R., J. W. Sanders, W. E. Dietrich, and S. D. Glaser, 2009: Influence of rock mass strength on the erosion rate of alpine cliffs. *Earth Surface Processes and Landforms*, **34** (10), 1339–1352.
- Mostovoy, G. V., R. L. King, K. R. Reddy, V. G. Kakani, and M. G. Filippova, 2006: Statistical Estimation of Daily Maximum and Minimum Air Temperatures from MODIS LST Data over the State of Mississippi. *GIScience & Remote Sensing*, **43** (1), 78–110.

- Mountrakis, G., J. Im, and C. Ogole, 2011: Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, **66** (3), 247–259.
- Mucina, L. and M. C. Rutherford, 2006: *The vegetation of South Africa, Lesotho and Swaziland*. South African National Biodiversity Institute.
- Munyati, C., P. Shaker, and M. Phasha, 2011: Using remotely sensed imagery to monitor savanna rangeland deterioration through woody plant proliferation: a case study from communal and biodiversity conservation rangeland sites in Mokopane, South Africa. *Environmental Monitoring and Assessment*, **176** (1–4), 293–311.
- Murton, J. B., R. Peterson, and J.-C. Ozouf, 2006: Bedrock Fracture by Ice Segregation in Cold Regions. *Science*, **314** (5802), 1127–1129.
- Mutanga, O. and A. K. Skidmore, 2004: Hyperspectral band depth analysis for a better estimation of grass biomass (*Cenchrus ciliaris*) measured under controlled laboratory conditions. *International Journal of Applied Earth Observation and Geoinformation*, **5** (2), 87–96.
- Mutanga, O. and A. K. Skidmore, 2007: Red edge shift and biochemical content in grass canopies. *ISPRS Journal of Photogrammetry and Remote Sensing*, **62** (1), 34–42.
- National Soil Survey Center NRCS USDA, 2011: *Field book for describing and sampling soils*. 3d ed., U.S. Department of Agriculture, Lincoln, Nebraska.
- Natural Resource Conservation Service (NRCS), 2013: Whitman county, WA soil survey. URL <http://websoilsurvey.sc.egov.usda.gov/App/HomePage.htm>.
- Nauss, T. and A. A. Kokhanovsky, 2006: Discriminating raining from non-raining clouds at mid-latitudes using multispectral satellite data. *Atmospheric Chemistry and Physics*, **6** (12), 5031–5036.
- Neteler, M., 2010: Estimating Daily Land Surface Temperatures in Mountainous Environments by Reconstructed MODIS LST Data. *Remote Sensing*, **2**, 333–351.
- Noetzli, J. and S. Gruber, 2009: Transient thermal effects in Alpine permafrost. *The Cryosphere*, **3** (1), 85–99.
- Noetzli, J., M. Hoelzle, and W. Haeberli, 2003: Mountain permafrost and recent Alpine rockfall events: a GIS-based approach to determine critical factors. *Proceedings of the 8th International Conference on Permafrost, Zürich.*, M. Phillips, S. Springman, and L. Arenson, Eds., 827–832.
- Nogués-Bravo, D., M. Araújo, M. Errea, and J. Martínez-Rica, 2007: Exposure of global mountain systems to climate warming during the 21st Century. *Global Environmental Change*, **17** (3–4), 420–428.
- Nordgren, A., 1988: Apparatus for the continuous, long-term monitoring of soil respiration rate in large numbers of samples. *Soil Biology and Biochemistry*, **20** (6), 955–957.

- Numata, I., D. A. Roberts, O. A. Chadwick, J. P. Schimel, L. S. Galvao, and J. V. Soares, 2008: Evaluation of hyperspectral data for pasture estimate in the Brazilian Amazon using field and imaging spectrometers. *Remote Sensing of Environment*, **112** (4), 1569–1583.
- Nyenhuis, M., 2006: Permafrost und Sedimenthaushalt in einem alpinen Geosystem. *Bonner Geographische Abhandlungen*, **116**, 142.
- Nyenhuis, M., M. Hoelzle, and R. Dikau, 2005: Rock glacier mapping and permafrost distribution modelling in the Turtmanntal, Valais, Switzerland. *Zeitschrift für Geomorphologie, NF*, **49** (3), 275–292.
- Obanawa, H., Y. S. Hayakawa, and Y. Matsukura, 2009: Rates of Slope Decline, Talus Growth and Cliff Retreat along the Shomyo River in Central Japan: A Space—Time Substitution Approach. *Geografiska Annaler. Series A, Physical Geography*, **91** (4), 269–278.
- Odeh, I., A. McBratney, and D. Chittleborough, 1995: Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. *Geoderma*, **67** (3), 215–226.
- Ohlmacher, G. C. and J. C. Davis, 2003: Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansas, USA. *Engineering Geology*, **69** (3–4), 331–343.
- Orchard, V. A. and F. Cook, 1983: Relationship between soil respiration and soil moisture. *Soil Biology and Biochemistry*, **15** (4), 447–453.
- Otto, J. and R. Dikau, 2004: Geomorphologic system analysis of a high mountain valley in the Swiss Alps. *Z. Geomorphol. N.F.*, **48**, 323–341.
- Otto, J.-C., K. Kleinod, O. König, M. Krautblatter, M. Nyenhuis, I. Roer, M. Schneider, B. Schreiner, and R. Dikau, 2007: HRSC-A data: a new high-resolution data set with multipurpose applications in physical geography. *Progress in Physical Geography*, **31** (2), 179–197.
- Otto, J.-C., L. Schrott, M. Jaboyedoff, and R. Dikau, 2009: Quantifying sediment storage in a high alpine valley (Turtmanntal, Switzerland). *Earth Surface Processes and Landforms*, **34** (13), 1726–1742.
- Otukei, J. and T. Blaschke, 2010: Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms. *International Journal of Applied Earth Observation and Geoinformation*, **12**, Supplement 1, S27–S31.
- Paine, A. D., 1985: 'Ergodic' reasoning in geomorphology. *Progress in Physical Geography*, **9** (1), 1–15.
- Pal, M., 2005: Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, **26** (1), 217–222.
- Palmborg, C. and A. Nordgren, 1993: Modelling microbial activity and biomass in forest soil with substrate quality measured using near infrared reflectance spectroscopy. *Soil Biology and Biochemistry*, **25** (12), 1713–1718.

- Panchal, G., A. Ganatra, Y. P. Kosta, and D. Panchal, 2011: Behaviour Analysis of Multilayer Perceptrons with Multiple Hidden Neurons and Hidden Layers. *International Journal of Computer Theory and Engineering*, **3** (2), 332–337.
- Paola, J. and R. Schowengerdt, 1995: A review and analysis of backpropagation neural networks for classification of remotely-sensed multi-spectral imagery. *International Journal of Remote Sensing*, **16** (16), 3033–3058.
- Pebesma, E., 2012: spacetime: Spatio-Temporal Data in R. *Journal of Statistical Software*, **51** (1), 1–30.
- Pebesma, E. and R. Bivand, 2013: *sp: classes and methods for spatial data*. URL <http://CRAN.R-project.org/package=sp>, R package version 1.0-5.
- Pebesma, E. and B. Gräler, 2013: *gstat: spatial and spatio-temporal geostatistical modelling, prediction and simulation*. URL <http://CRAN.R-project.org/package=gstat>, R package version 1.0-16.
- Pebesma, E. J., 2004: Multivariable geostatistics in S: the gstat package. *Computers & Geosciences*, **30** (7), 683–691.
- Peirce, C., 1902: Verification. *Dictionary of Philosophy and Psychology*, J. Baldwin, Ed., Macmillan, New York, 761–762.
- Phillips, J., 1988: The role of spatial scale in geomorphic systems. *Geographical Analysis*, **20**, 308–317.
- Phillips, J. D., 2003: Sources of nonlinearity and complexity in geomorphic systems. *Progress in Physical Geography*, **27** (1), 1–23.
- Pickett, S. T. A., 1989: *Space-for-Time Substitution as an Alternative to Long-Term Studies*, 110–135. Springer New York, New York, NY.
- Pierce, F. and T. Elliott, 2008: Regional and on-farm wireless sensor networks for agricultural systems in Eastern Washington. *Computers and Electronics in Agriculture*, **61** (1), 32–43.
- Pinheiro, J. and D. Bates, 2009: *Mixed-Effects Models in S and S-PLUS*. Statistics and Computing, Springer.
- Porter, J., P. Arzberger, H.-W. Braun, P. Bryant, S. Gage, T. Hansen, P. Hanson, C.-C. Lin, F.-P. Lin, T. Kratz, W. Michener, S. Shapiro, and T. Williams, 2005: Wireless Sensor Networks for Ecology. *Bioscience*, **55** (7), 561.
- Porter, J. and D. Demeritt, 2012: Flood-Risk Management, Mapping, and Planning: The Institutional Politics of Decision Support in England. *Environment and Planning A*, **44** (10), 2359–2378.
- Potter, C. S., J. T. Randerson, C. B. Field, P. A. Matson, P. M. Vitousek, H. A. Mooney, and S. A. Klooster, 1993: Terrestrial ecosystem production: A process model based on global satellite and surface data. *Global Biogeochemical Cycles*, **7** (4), 811–841.
- Prick, A., 1997: Critical Degree of Saturation as a Threshold Moisture Level in Frost Weathering of Limestones. *Permafrost and Periglacial Processes*, **8** (1), 91–99.

- Prick, A., 2003: Frost weathering and rock fall in an arctic environment, Longyearbyen, Svalbard. *Proceedings of the Eighth International Conference on Permafrost*. Balkema, Lisse, M. Phillips, Ed., 907–912.
- Prigent, C., 2010: Precipitation retrieval from space: An overview. *Comptes Rendus Geoscience*, **342**, 380–389.
- Psomas, A., 2008: Hyperspectral remote sensing for ecological analyses of grassland ecosystems - spectral separability and derivation of NPP-related biophysical and biochemical parameters. Ph.D. thesis, University of Zürich (UNIZH), Zürich.
- Psomas, A., M. Kneubuhler, S. Huber, K. Itten, and N. E. Zimmermann, 2011: Hyperspectral remote sensing for estimating aboveground biomass and for exploring species richness patterns of grassland habitats. *International Journal of Remote Sensing*, **32** (24), 9007–9031.
- Qiu, H., D. Huggins, J. Wu, M. Barber, D. Mccool, and S. Dun, 2011: Residue management impacts on field-scale snow distribution and soil water storage. *Transactions of the ASABE*, **54**, 1639–1647.
- R Core Team, 2013: *R: A Language and Environment for Statistical Computing*. Vienna, Austria, R Foundation for Statistical Computing, URL <http://www.R-project.org/>.
- R Core Team, 2014: *R: A Language and Environment for Statistical Computing*. Vienna, Austria, R Foundation for Statistical Computing, URL <http://www.R-project.org/>.
- R Core Team, 2015: *R: A Language and Environment for Statistical Computing*. Vienna, Austria, R Foundation for Statistical Computing, URL <https://www.R-project.org/>.
- R Core Team, 2016: *R: A Language and Environment for Statistical Computing*. Vienna, Austria, R Foundation for Statistical Computing, URL <https://www.R-project.org/>.
- Raich, J. W., E. B. Rastetter, J. M. Melillo, D. W. Kicklighter, P. A. Steudler, B. J. Peterson, A. L. Grace, B. Moore, and C. J. Vorosmarty, 1991: Potential Net Primary Productivity in South America: Application of a Global Model. *Ecological Applications*, **1** (4), 399–429.
- Rapp, A., 1960: Talus Slopes and Mountain Walls at Tempelfjorden, Spitsbergen. A geomorphological Study of the Denudation of Slopes in an Arctic Locality. *Oslo University Press, Uppsala*, 96.
- Rasouli, K., W. W. Hsieh, and A. J. Cannon, 2012: Daily streamflow forecasting by machine learning methods with weather and climate inputs. *Journal of Hydrology*, **414–415**, 284–293.
- Ravel, L. and P. Deline, 2011: Climate influence on rockfalls in high-Alpine steep rockwalls: The north side of the Aiguilles de Chamonix (Mont Blanc massif) since the end of the ‘Little Ice Age’. *The Holocene*, **21** (2), 357–365.

- Ravanel, L. and P. Deline, 2015: *Rockfall Hazard in the Mont Blanc Massif Increased by the Current Atmospheric Warming*, 425–428. Springer International Publishing, Cham.
- Reeves, J. B., 2010: Near- versus mid-infrared diffuse reflectance spectroscopy for soil analysis emphasizing carbon and laboratory versus on-site analysis: Where are we and what needs to be done? *Geoderma*, **158** (1), 3–14.
- Reeves, J. B., G. W. McCarty, and V. B. Reeves, 2001: Mid-infrared Diffuse Reflectance Spectroscopy for the Quantitative Analysis of Agricultural Soils. *Journal of Agricultural and Food Chemistry*, **49** (2), 766–772.
- Reichenbach, P., M. Galli, M. Cardinali, F. Guzzetti, and F. Ardizzone, 2005: *Geomorphological Mapping to Assess Landslide Risk: Concepts, Methods and Applications in the Umbria Region of Central Italy*, 429–468. John Wiley & Sons, Ltd.
- Reichstein, M., *et al.*, 2003: Modeling temporal and large-scale spatial variability of soil respiration from soil water availability, temperature and vegetation productivity indices. *Global Biogeochemical Cycles*, **17** (4).
- Revolution Analytics and S. Weston, 2014: *doParallel: Foreach parallel adaptor for the parallel package*. URL <http://CRAN.R-project.org/package=doParallel>, R package version 1.0.8.
- Rhee, J. and J. Im, 2014: Estimating High Spatial Resolution Air Temperature for Regions with Limited in situ Data Using MODIS Products. *Remote Sensing*, **6** (8), 7360.
- Ridgeway, G., 2015: *gbm: Generalized Boosted Regression Models*. URL <https://CRAN.R-project.org/package=gbm>, R package version 2.1.1.
- Ripley, B. and W. Venables, 2016: *nnet: Feed-Forward Neural Networks and Multinomial Log-Linear Models*. URL <http://CRAN.R-project.org/package=nnet>, R package version 7.3-12.
- Rivolta, G., F. S. Marzano, E. Coppola, and M. Verdecchia, 2006: Artificial neural-network technique for precipitation nowcasting from satellite imagery. *Advances in Geosciences*, **7**, 97–103.
- Roberts, D. R., V. Bahn, S. Ciuti, M. S. Boyce, J. Elith, G. Guillera-Arroita, S. Hauenstein, J. J. Lahoz-Monfort, B. Schröder, W. Thuiller, D. I. Warton, B. A. Wintle, F. Hartig, and C. F. Dormann, 2017: Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, doi:10.1111/ecog.02881.
- Rodriguez-Galiano, V., B. Ghimire, J. Rogan, M. Chica-Olmo, and J. Rigol-Sanchez, 2012: An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, **67**, 93–104.
- Roebeling, R. and I. Holleman, 2009: SEVIRI rainfall retrieval and validation using weather radar observations. *Journal of Geophysical Research*, **114**, D21 202.

- Roebeling, R. A., A. J. Feijt, and P. Stammes, 2006: Cloud property retrievals for climate monitoring: Implications of differences between Spinning Enhanced Visible and Infrared Imager (SEVIRI) on METEOSAT-8 and Advanced Very High Resolution Radiometer (AVHRR) on NOAA-17. *Journal of Geophysical Research: Atmospheres*, **111** (D20).
- Roques, K. G., T. G. O'Connor, and A. R. Watkinson, 2001: Dynamics of Shrub Encroachment in an African Savanna: Relative Influences of Fire, Herbivory, Rainfall and Density Dependence. *Journal of Applied Ecology*, **38** (2), 268–280.
- Rosenfeld, D. and I. M. Lensky, 1998: Satellite-Based Insights into Precipitation Formation Processes in Continental and Maritime Convective Clouds. *Bulletin of the American Meteorological Society*, **79** (11), 2457–2476.
- Rosenfield, G. H. and K. Fitzpatrick-Lins, 1986: A coefficient of agreement as a measure of thematic classification accuracy. *Photogrammetric Engineering & Remote Sensing*, **52** (2), 223–227.
- Rumpel, C., L. Janik, J. Skjemstad, and I. Kögel-Knabner, 2001: Quantification of carbon derived from lignite in soils using mid-infrared spectroscopy and partial least squares. *Organic Geochemistry*, **32** (6), 831–839.
- Ruß, G. and A. Brenning, 2010: *Data Mining in Precision Agriculture: Management of Spatial Information*, 350–359. Springer Berlin Heidelberg.
- Sandmeier, S., C. Müller, B. Hosgood, and G. Andreoli, 1998: Physical Mechanisms in Hyperspectral BRDF Data of Grass and Watercress. *Remote Sensing of Environment*, **66** (2), 222–233.
- Santi, P. M., C. P. Russell, J. D. Higgins, and J. I. Spriet, 2009: Modification and statistical analysis of the Colorado Rockfall Hazard Rating System. *Engineering Geology*, **104** (1–2), 55–65.
- Sarkar, D., 2008: *Lattice: multivariate data visualization with R*. Springer.
- Sass, O., 1998: Die Steuerung von Steinschlagmenge durch Mikroklima, Gesteinsfeuchte und Gesteineigenschaften im westlichen Karwendelgebirge (Bayerische Alpen). *Münchner Geographische Abhandlungen. Reihe B. Geobuch-Verlag, München*, 175.
- Sass, O., 2004: Rock Moisture Fluctuations During Freeze-thaw Cycles: Preliminary Results from Electrical Resistivity Measurements. *Polar Geography*, **28** (1), 13–31.
- Sass, O., 2005a: Rock moisture measurements: techniques, results, and implications for weathering. *Earth Surface Processes and Landforms*, **30** (3), 359–374.
- Sass, O., 2005b: Spatial patterns of rockfall intensity in the northern Alps. *Z. Geomorphol. N.F.*, **138**, 51–65.
- Sass, O., 2005c: Temporal Variability of Rockfall in the Bavarian Alps, Germany. *Arctic, Antarctic, and Alpine Research*, **37** (4), 564–573.
- Sass, O., 2010: Spatial and temporal patterns of talus activity – a lichenometric approach in the Stubai Alps, Austria. *Geografiska Annaler: Series A, Physical Geography*, **92** (3), 375–391.

- Schlunegger, F. and M. Hinderer, 2001: Crustal uplift in the Alps: why the drainage pattern matters. *Terra Nova*, **13** (6), 425–432.
- Schlüchter, C., 2004: The Swiss glacial record – a schematic summary. *Quaternary Glaciations Extent and Chronology Part I: Europe*, J. Ehlers and P. Gibbard, Eds., Elsevier, Developments in Quaternary Sciences, Vol. 2, Part 1, 413 – 418.
- Schneider, D. P. and D. B. Reusch, 2016: Antarctic and Southern Ocean Surface Temperatures in CMIP5 Models in the Context of the Surface Energy Budget. *Journal of Climate*, **29** (5), 1689–1716.
- Schrott, L. and T. Hoffmann, 2003: Determining sediment thickness of talus slopes and valley fill deposits using seismic refraction — a comparison of 2D interpretation tools. *Zeitschrift für Geomorphologie N.F.*, **132**, 71–87.
- Selby, M., 1980: A rock mass strength classification for geomorphic purposes: with tests from Antarctica and New Zealand. *Zeitschrift für Geomorphologie*, **24**, 31–51.
- Selby, M., (Ed.) , 1982a: *Hillslope Materials and Processes*. Oxford University Press, Oxford.
- Selby, M. J., 1982b: Controls on the stability and inclinations of hillslopes formed on hard rock. *Earth Surface Processes and Landforms*, **7** (5), 449–467.
- Seybold, C. A., D. S. Harms, M. Balks, J. Aislabie, R. F. Paetzold, J. Kimble, and R. Sletten, 2009: Soil Climate Monitoring Project in the Ross Island Region of Antarctica. *Soil Survey Horizons*, 52–57.
- Sheehy, D., D. Miller, and D. Johnson, 2006: Transformation of traditional pastoral livestock systems on the Tibetan steppe. *Sécheresse*, **17** (1), 142–151.
- Shen, C., C. Li, and Y. Si, 2016: Spatio-temporal autocorrelation measures for nonstationary series: A new temporally detrended spatio-temporal Moran's index. *Physics Letters A*, **380** (1), 106–116.
- Shen, M., Y. Tang, J. Klein, P. Zhang, S. Gu, A. Shimono, and J. Chen, 2008: Estimation of aboveground biomass using in situ hyperspectral measurements in five major grassland ecosystems on the Tibetan Plateau. *Journal of Plant Ecology*, **1** (4), 247–257.
- Shi, L., P. Liu, I. Kloog, M. Lee, A. Kosheleva, and J. Schwartz, 2016: Estimating daily air temperature across the Southeastern United States using high-resolution satellite data: A statistical modeling study. *Environmental Research*, **146**, 51–58.
- Shi, Y., L. Song, Z. Xia, Y. Lin, R. B. Myneni, S. Choi, L. Wang, X. Ni, C. Lao, and F. Yang, 2015: Mapping Annual Precipitation across Mainland China in the Period 2001–2010 from TRMM3B43 Product Using Spatial Downscaling Approach. *Remote Sensing*, **7** (5), 5849–5878.
- Shirzadi, A., L. Saro, O. Hyun Joo, and K. Chapi, 2012: A GIS-based logistic regression model in rock-fall susceptibility mapping along a mountainous road: Salavat Abad case study, Kurdistan, Iran. *Natural Hazards*, **64** (2), 1639–1656.

- Shmueli, G., 2010: To Explain or to Predict? *Statistical Science*, **25** (3), 289–310.
- Shortridge, J. E., S. D. Guikema, and B. F. Zaitchik, 2016: Machine learning methods for empirical streamflow simulation: a comparison of model accuracy, interpretability, and uncertainty in seasonal watersheds. *Hydrology and Earth System Sciences*, **20** (7), 2611–2628.
- Simmer, C., *et al.*, 2015: Monitoring and Modeling the Terrestrial System from Pores to Catchments: The Transregional Collaborative Research Center on Patterns in the Soil–Vegetation–Atmosphere System. *Bulletin of the American Meteorological Society*, **96** (10), 1765–1787.
- Skarpe, C., 1991: Spatial patterns and dynamics of woody vegetation in an arid savanna. *Journal of Vegetation Science*, **2** (4), 565–572.
- Skofronick-Jackson, G., W. A. Petersen, W. Berg, C. Kidd, E. F. Stocker, D. B. Kirschbaum, R. Kakar, S. A. Braun, G. J. Huffman, T. Iguchi, P. E. Kirstetter, C. Kummerow, R. Meneghini, R. Oki, W. S. Olson, Y. N. Takayabu, K. Furukawa, and T. Wilheit, 2017: The Global Precipitation Measurement (GPM) Mission for Science and Society. *Bulletin of the American Meteorological Society*, doi:10.1175/BAMS-D-15-00306.1.
- Smith, M., P. Paron, and J. Griffiths, (Eds.) , 2011: *Geomorphological Mapping - Methods and Applications*, Vol. 15. Elsevier Science.
- Snepvangers, J., G. Heuvelink, and J. Huisman, 2003: Soil water content interpolation using spatio-temporal kriging with external drift. *Geoderma*, **112** (3–4), 253–271.
- Sohrabinia, M., P. Zawar-Reza, and W. Rack, 2014: Spatio-temporal analysis of the relationship between LST from MODIS and air temperature in New Zealand. *Theoretical and Applied Climatology*, **119** (3), 567–583.
- Srivastava, P. K., D. Han, M. R. Ramirez, and T. Islam, 2013: Machine Learning Techniques for Downscaling SMOS Satellite Soil Moisture Using MODIS Land Surface Temperature for Hydrological Application. *Water Resources Management*, **27** (8), 3127–3144.
- Stahl, K., R. Moore, J. Floyer, M. Asplin, and I. McKendry, 2006: Comparison of approaches for spatial interpolation of daily air temperature in a large region with complex topography and highly variable station density. *Agricultural and Forest Meteorology*, **139** (3–4), 224–236.
- Statnikov, A., L. Wang, and C. F. Aliferis, 2008: A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification. *BMC Bioinformatics*, **9**, 319–319.
- Steele, B., 2000: Combining multiple classifiers: an application using spatial and remotely sensed information for land cover type mapping. *Remote Sensing of Environment*, **74** (3), 545–556.
- Stengel, M., A. Kniffka, J. F. Meirink, M. Lockhoff, J. Tan, and R. Hollmann, 2014: CLAAS: the CM SAF cloud property data set using SEVIRI. *Atmospheric Chemistry and Physics*, **14** (8), 4297–4311.

- Sternberg, T., 2012: Piospheres and Pastoralists: Vegetation and Degradation in Steppe Grasslands. *Human Ecology*, **40** (6), 811–820.
- Stevens, A., M. Nocita, G. Tóth, L. Montanarella, and B. van Wesemael, 2013: Prediction of Soil Organic Carbon at the European Scale by Visible and Near InfraRed Reflectance Spectroscopy. *PLOS ONE*, **8** (6), 1–13.
- Stumpf, A. and N. Kerle, 2011: Object-oriented mapping of landslides using Random Forests. *Remote Sensing of Environment*, **115** (10), 2564–2577.
- Sun, J., G. W. Cheng, W. P. Li, Y. K. Sha, and Y. C. Yang, 2013: On the Variation of NDVI with the Principal Climatic Elements in the Tibetan Plateau. *Remote Sensing*, **5** (4), 1894–1911.
- Swets, J., 1988: Measuring the accuracy of diagnostic systems. *Science*, **240** (4857), 1285–1293.
- Symeonakis, E. and T. P. Higginbottom, 2014: Bush encroachment monitoring using multi-temporal Landsat data and random forests. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XL-2.
- Tanarro, L. M. and J. Muñoz, 2012: Rockfalls in the Duratón canyon, central Spain: Inventory and statistical analysis. *Geomorphology*, **169–170**, 17–29.
- Tapakis, R. and A. Charalambides, 2013: Equipment and methodologies for cloud detection and classification: A review. *Solar Energy*, **95**, 392–430.
- Tapiador, F. J., C. Kidd, K.-L. Hsu, and F. Marzano, 2004: Neural networks in satellite rainfall estimation. *Meteorological Applications*, **11** (1), 83–91.
- Tebbi, M. and B. Haddad, 2016: Artificial intelligence systems for rainy areas detection and convective cells' delineation for the south shore of Mediterranean Sea during day and nighttime using MSG satellite images. *Atmospheric Research*, **178–179**, 380–392.
- Terzaghi, K., 1962: Stability of Steep Slopes on Hard Unweathered Rock. *Géotechnique*, **12** (4), 251–270.
- Thenkabail, P. S., R. B. Smith, and E. De Pauw, 2002: Evaluation of narrowband and broadband vegetation indices for determining optimal hyperspectral wavebands for agricultural crop characterization. *Photogrammetric Engineering & Remote Sensing*, **68** (6), 607–621.
- Thies, B. and J. Bendix, 2011: Satellite based remote sensing of weather and climate: recent achievements and future perspectives. *Meteorological Applications*, **18** (3), 262–295.
- Thies, B., H. Meyer, T. Nauss, and J. Bendix, 2014: Projecting land-use and land-cover changes in a tropical mountain forest of Southern Ecuador. *Journal of Land Use Science*, **9** (1), 1–33.
- Thies, B., T. Nauss, and J. Bendix, 2008a: Discriminating raining from non-raining cloud areas at mid-latitudes using meteosat second generation SEVIRI night-time data. *Meteorological Applications*, **15** (2), 219–230.

- Thies, B., T. Nauss, and J. Bendix, 2008b: Discriminating raining from non-raining clouds at mid-latitudes using meteosat second generation daytime data. *Atmospheric Chemistry and Physics*, **8** (9), 2341–2349.
- Thies, B., T. Nauss, and J. Bendix, 2008c: Precipitation process and rainfall intensity differentiation using Meteosat Second Generation Spinning Enhanced Visible and Infrared Imager data. *Journal of Geophysical Research: Atmospheres*, **113** (D23).
- Thomson, B. C., N. J. Ostle, N. P. McNamara, A. S. Whiteley, and R. I. Griffiths, 2010: Effects of sieving, drying and rewetting upon soil bacterial community structure and respiration rates. *Journal of Microbiological Methods*, **83** (1), 69–73.
- Tian, B., M. A. Shaikh, M. R. Azimi-Sadjadi, T. H. V. Haar, and D. L. Reinke, 1999: A study of cloud classification with neural networks using spectral and textural features. *IEEE Transactions on Neural Networks*, **10** (1), 138–151.
- Tobler, W. R., 1970: A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, **46**, 234–240.
- Tricart, J. and A. Cailleux, (Eds.) , 1972: *Introduction to Climatic Geomorphology*. Longman, London.
- Tripathi, S., V. Srinivas, and R. S. Nanjundiah, 2006: Downscaling of precipitation for climate change scenarios: A support vector machine approach. *Journal of Hydrology*, **330** (3–4), 621–640.
- Tucker, C. J., 1979: Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, **8** (2), 127–150.
- Uddstrom, M. J. and W. R. Gray, 1996: Satellite Cloud Classification and Rain-Rate Estimation Using Multispectral Radiances and Measures of Spatial Texture. *Journal of Applied Meteorology*, **35** (6), 839–858.
- U.S Geological Survey, 2007: *Landsat Image Mosaic of Antarctica (LIMA): U.S Geological Survey Fact Sheet 2007 - 3116*.
- Vaglio Laurin, G., J. C.-W. Chan, Q. Chen, J. A. Lindsell, D. A. Coomes, L. Guerriero, F. D. Frate, F. Miglietta, and R. Valentini, 2014: Biodiversity Mapping in a Tropical West African Forest with Airborne Hyperspectral Data. *PLOS ONE*, **9** (6), e97910.
- Valentine, A. and L. Kalnins, 2016: An introduction to learning algorithms and potential applications in geomorphometry and Earth surface dynamics. *Earth Surface Dynamics*, **4** (2), 445–460.
- van Tatenhove, F. and R. Dikau, 1990: Past and Present Permafrost Distribution in the Turtmantal, Wallis, Swiss Alps. *Arctic and Alpine Research*, **22** (3), 302–316.
- van Zyl, T., 2014: Machine Learning on Geospatial Big Data. *Big Data*, CRC Press, 133–148.

- Vanacker, V., M. Vanderschaeghe, G. Govers, E. Willems, J. Poesen, J. Deckers, and B. D. Bievre, 2003: Linking hydrological, infinite slope stability and land-use change models through GIS for assessing the impact of deforestation on slope stability in high Andean watersheds. *Geomorphology*, **52** (3–4), 299–315.
- Vancutsem, C., P. Ceccato, T. Dinku, and S. J. Connor, 2010: Evaluation of MODIS land surface temperature data to estimate air temperature in different ecosystems over Africa. *Remote Sensing of Environment*, **114** (2), 449–465.
- Vanhala, P., K. Karhu, M. Tuomi, K. Björklöf, H. Fritze, and J. Liski, 2008: Temperature sensitivity of soil organic matter decomposition in southern and northern areas of the boreal forest zone. *Soil Biology and Biochemistry*, **40** (7), 1758–1764.
- Van't Hoff, J., 1898: *Lectures on Theoretical and Physical Chemistry. Part I. Chemical Dynamics* (translated by R. A. Lehfeldt). Edward Arnold, London.
- Vanwalleghe, T., M. V. D. Eeckhaut, J. Poesen, G. Govers, and J. Deckers, 2008: Spatial analysis of factors controlling the presence of closed depressions and gullies under forest: Application of rare event logistic regression. *Geomorphology*, **95** (3–4), 504–517.
- Vapnik, V. N., 1995: *The Nature of Statistical Learning Theory*. 2d ed., Springer, New York.
- Varnes, D., (Ed.) , 1984: *Landslide Hazard Zonation: A Review of Principles and Practice*. UNESCO, Paris.
- Vehling, L., J. Rohn, and M. Moser, 2016: Quantification of small magnitude rockfall processes at a proglacial high mountain site, Gepatsch glacier (Tyrol, Austria). *Zeitschrift für Geomorphologie, Supplementary Issues*, **60** (1), 93–108.
- Veldkamp, A. and E. Lambin, 2001: Predicting land-use change. *Agriculture, Ecosystems and Environment*, **85**, 1–6.
- Venables, W. N. and B. D. Ripley, 2002: *Modern Applied Statistics with S*. 4th ed., Springer, New York.
- Vereecken, H., Y. Pachepsky, C. Simmer, J. Rihani, A. Kunoth, W. Korres, A. Graf, H.-H. Franssen, I. Thiele-Eich, and Y. Shao, 2016: On the role of patterns in understanding the functioning of soil-vegetation-atmosphere systems. *Journal of Hydrology*, **542**, 63–86.
- Verleysdonk, S., M. Krautblatter, and R. Dikau, 2011: Sensitivity and path dependence of mountain permafrost systems. *Geografiska Annaler: Series A, Physical Geography*, **93** (2), 113–135.
- Veronesi, F., R. Corstanje, and T. Mayr, 2012: Mapping soil compaction in 3D with depth functions. *Soil and Tillage Research*, **124**, 111–118.
- Verrelst, J., J. Muñoz, L. Alonso, J. Delegido, J. P. Rivera, G. Camps-Valls, and J. Moreno, 2012: Machine learning regression algorithms for biophysical parameter retrieval: Opportunities for Sentinel-2 and -3. *Remote Sensing of Environment*, **118**, 127–139.

- Vicente, G. A., J. C. Davenport, and R. A. Scofield, 2002: The role of orographic and parallax corrections on real time high resolution satellite rainfall rate distribution. *International Journal of Remote Sensing*, **23** (2), 221–230.
- Vicente, G. a., R. a. Scofield, W. P. Menzel, and G. A. Vincente, 1998: The Operational GOES Infrared Rainfall Estimation Technique. *Bulletin of the American Meteorological Society*, **79** (9), 1883–1898.
- Viles, H. A., 2013: Linking weathering and rock slope instability: non-linear perspectives. *Earth Surface Processes and Landforms*, **38** (1), 62–70.
- Viscarra Rossel, R. A., H. J. Taylor, and A. B. McBratney, 2007: Multivariate calibration of hyperspectral γ -ray energy spectra for proximal soil sensing. *European Journal of Soil Science*, **58** (1), 343–353.
- Vogt, J. V., A. A. Viau, and F. Paquet, 1997: Mapping regional air temperature fields using satellite-derived surface skin temperatures. *International Journal of Climatology*, **17** (14), 1559–1579.
- Vorpahl, P., H. Elsenbeer, M. Märker, and B. Schröder, 2012: How can statistical models help to determine driving factors of landslides? *Ecological Modelling*, **239**, 27–39.
- Vuolo, F., N. Neugebauer, S. F. Bolognesi, C. Atzberger, and G. D’Urso, 2013: Estimation of Leaf Area Index Using DEIMOS-1 Data: Application and Transferability of a Semi-Empirical Relationship between two Agricultural Areas. *Remote Sensing*, **5** (3), 1274.
- Walder, J. and B. Hallet, 1985: A theoretical model of the fracture of rock during freezing. *Geological Society of America Bulletin*, **96** (3), 336–346.
- Wan, Z., 2008: New refinements and validation of the MODIS Land-Surface Temperature/Emissivity products. *Remote Sensing of Environment*, **112** (1), 59–74.
- Wang, J., B. Fu, Y. Qiu, L. Chen, and Z. Wang, 2001: Geostatistical analysis of soil moisture variability on Da Nangou catchment of the loess plateau, China. *Environmental Geology*, **41** (1), 113–120.
- Wang, W., R. Dalal, P. Moody, and C. Smith, 2003: Relationships of soil respiration to microbial biomass, substrate availability and clay content. *Soil Biology and Biochemistry*, **35** (2), 273–284.
- Wang, Y. and S. Hou, 2009: A new interpolation method for Antarctic surface temperature. *Progress in Natural Science*, **19** (12), 1843–1849.
- Wang, Y., M. Wang, and J. Zhao, 2013: A Comparison of MODIS LST Retrievals with in Situ Observations from AWS over the Lambert Glacier Basin, East Antarctica. *International Journal of Geosciences*, **4**, 611–617.
- Ward, D., 2005: Do we understand the causes of bush encroachment in African savannas? *African Journal of Range and Forage Science*, **22**, 101–105.
- Warnes, G., B. Bolker, G. Gorjanc, G. Grothendieck, A. Korosec, T. Lumley, D. MacQueen, A. Magnusson, J. Rogers, and et al., 2014: *gdata: Various R programming tools for data manipulation*. URL <http://CRAN.R-project.org/package=gdata>, R package version 2.13.3.

- Waske, B., J. A. Benediktsson, K. Árnason, and J. R. Sveinsson, 2009: Mapping of hyperspectral AVIRIS data using machine-learning algorithms. *Canadian Journal of Remote Sensing*, **35** (sup1), S106–S116.
- Welch, R. M., S. K. Sengupta, and D. W. Chen, 1988: Cloud field classification based upon high spatial resolution textural features: 1. Gray level co-occurrence matrix approach. *Journal of Geophysical Research: Atmospheres*, **93** (D10), 12 663–12 681.
- Wessels, K., S. Prince, and I. Reshef, 2008: Mapping land degradation by comparison of vegetation production to spatially derived estimates of potential production. *Journal of Arid Environments*, **72** (10), 1940–1949.
- Westermann, S., M. Langer, and J. Boike, 2012: Systematic bias of average winter-time land surface temperatures inferred from MODIS at a site on Svalbard, Norway. *Remote Sensing of Environment*, **118**, 162–167.
- Western Regional Climate Center, 2013: Climate summary, Pullman, WA. URL <http://www.wrcc.dri.edu>.
- Whalley, W., 1984: Rockfalls. *Slope Instability*, D. Brunsten and D. Prior, Eds., Wiley, Chichester, 217–256.
- Wichmann, V., T. Heckmann, F. Haas, and M. Becht, 2009: A new modelling approach to delineate the spatial extent of alpine sediment cascades. *Geomorphology*, **111** (1–2), 70–78.
- Wickham, H., 2014: *plyr: Tools for splitting, applying and combining data*. URL <http://CRAN.R-project.org/package=gstat>, R package version 1.8.1.
- Wiegand, K., D. Ward, and D. Saltz, 2005: Multi-scale patterns and bush encroachment in an arid savanna with a shallow soil layer. *Journal of Vegetation Science*, **16** (3), 311–320.
- Wilks, D., 2006: *Statistical Methods in the Atmospheric Sciences*. Academic Press, Burlington.
- Wilson, D. J., A. W. Western, R. B. Grayson, A. A. Berg, M. S. Lear, M. Rodell, J. S. Famiglietti, R. A. Woods, and T. A. McMahon, 2003: Spatial distribution of soil moisture over 6 and 30 cm depth, Mahurangi river catchment, New Zealand. *Journal of Hydrology*, **276** (1–4), 254–274.
- Wöllauer, S., S. Forteva, and T. Nauss, 2015: On demand processing of climate station sensor data. *EGU General Assembly Conference Abstracts*, EGU General Assembly Conference Abstracts, Vol. 17, 5231.
- Wyllie, D. and C. Mah, (Eds.) , 2004: *Rock Slope Engineering: Civil and Mining*. Spon Press, London/New York.
- Xu, Y., A. Knudby, and H. C. Ho, 2014: Estimating daily maximum air temperature from MODIS in British Columbia, Canada. *International Journal of Remote Sensing*, **35** (24), 8108–8121.
- Xu, Y., Z. Qin, and Y. Shen, 2012: Study on the estimation of near-surface air temperature from MODIS data by statistical methods. *International Journal of Remote Sensing*, **33** (24), 7629–7643.

- Yang, Y., J. Fang, Y. Pan, and C. Ji, 2009: Aboveground biomass in Tibetan grasslands. *Journal of Arid Environments*, **73** (1), 91–95.
- Zacharias, S., *et al.*, 2011: A Network of Terrestrial Environmental Observatories in Germany. *Vadose Zone Journal*, **10** (3), 955–973.
- Zech, W., F. Ziegler, I. Kögel-Knabner, and L. Haumaier, 1992: Humic substances distribution and transformation in forest soils. *Science of The Total Environment*, **117**, 155–174.
- Zha, Y., J. Gao, S. Ni, Y. Liu, J. Jiang, and Y. Wei, 2003: A spectral reflectance-based approach to quantification of grassland cover from Landsat TM imagery. *Remote Sensing of Environment*, **87** (2-3), 371–375.
- Zhang, X., M. A. Friedl, C. B. Schaaf, A. H. Strahler, J. C. Hodges, F. Gao, B. C. Reed, and A. Huete, 2003: Monitoring vegetation phenology using MODIS. *Remote Sensing of Environment*, **84** (3), 471–475.
- Zheng, Z.-M., G.-R. Yu, Y.-L. Fu, Y.-S. Wang, X.-M. Sun, and Y.-H. Wang, 2009: Temperature sensitivity of soil respiration is affected by prevailing climatic conditions and soil organic carbon content: A trans-China based case study. *Soil Biology and Biochemistry*, **41** (7), 1531–1540.
- Zhong, L., Y. Ma, M. S. Salama, and Z. Su, 2010: Assessment of vegetation dynamics and their response to variations in precipitation and temperature in the Tibetan Plateau. *Climatic Change*, **103** (3-4), 519–535.
- Zhou, T., P. Shi, D. Hui, and Y. Luo, 2009: Global pattern of temperature sensitivity of soil heterotrophic respiration (Q₁₀) and its implications for carbon-climate feedback. *Journal of Geophysical Research: Biogeosciences*, **114** (G2).
- Zhu, W., A. Lü, and S. Jia, 2013: Estimation of daily maximum and minimum air temperature using MODIS land surface temperature products. *Remote Sensing of Environment*, **130**, 62–73.
- Zimmermann, M., J. Leifeld, and J. Fuhrer, 2007: Quantifying soil organic carbon fractions by infrared-spectroscopy. *Soil Biology and Biochemistry*, **39** (1), 224–231.
- Zvoleff, A., 2015: *glcm: Calculate textures from grey-level co-occurrence matrices (GLCMs) in R*. URL <http://CRAN.R-project.org/package=glcm>, R package version 1.2.

ERKLÄRUNG

Ich erkläre an Eides statt, dass ich meine Dissertation

*Data-driven model development in environmental geography -
Methodological advancements and scientific applications*

selbstständig und ohne unerlaubte Hilfe angefertigt und mich dabei keinerlei anderer als der von mir ausdrücklich bezeichneten Quellen und Hilfen bedient habe.

Die Dissertation wurde in der jetzigen oder einer ähnlichen Form noch an keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Marburg, 17.07.2017

Hanna Meyer