Remote Sensing of Environment 141 (2014) 129–143

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Improving the accuracy of rainfall rates from optical satellite sensors with machine learning — A random forests-based approach applied to MSG SEVIRI

Meike Kühnlein [a,*], Tim Appelhans [a], Boris Thies [b], Thomas Nauss [a]

[a] Environmental Informatics, Faculty of Geography, Philipps-University Marburg, Marburg, Germany
[b] Laboratory for Climatology and Remote Sensing, Faculty of Geography, Philipps-University Marburg, Marburg, Germany

ARTICLE INFO

Article history:
Received 25 July 2013
Received in revised form 26 October 2013
Accepted 26 October 2013
Available online 22 November 2013

Keywords:
Rainfall rate
Rainfall retrieval
Random forests
Machine learning
MSG SEVIRI
Geostationary satellites
Optical sensors

ABSTRACT

The present study aims to investigate the potential of the random forests ensemble classification and regression technique to improve rainfall rate assignment during day, night and twilight (resulting in 24-hour precipitation estimates) based on cloud physical properties retrieved from Meteosat Second Generation (MSG) Spinning Enhanced Visible and InfraRed Imager (SEVIRI) data.

Random forests (RF) models contain a combination of characteristics that make them well suited for its application in precipitation remote sensing. One of the key advantages is the ability to capture non-linear association of patterns between predictors and response which becomes important when dealing with complex non-linear events like precipitation. Due to the deficiencies of existing optical rainfall retrievals, the focus of this study is on assigning rainfall rates to precipitating cloud areas in connection with extra-tropical cyclones in mid-latitudes including both convective and advective-stratiform precipitating cloud areas. Hence, the rainfall rates are assigned to rain areas previously identified and classified according to the precipitation formation processes. As predictor variables water vapor-IR differences and IR cloud top temperature are used to incorporate information on cloud top height. $\Delta T_{8.7-10.8}$ and $\Delta T_{10.8-12.1}$ are considered to supply information about the cloud phase. Furthermore, spectral SEVIRI channels ($VIS_{0.6}$, $VIS_{0.8}$, $NIR_{1.6}$) and cloud properties (cloud effective radius, cloud optical thickness) are used to include information about the cloud water path during daytime, while suitable combinations of temperature differences ($\Delta T_{3.9-10.8}$, $\Delta T_{3.9-7.3}$) are considered during night-time.

The development of the rainfall rate retrieval technique is realised in three steps. First, an extensive tuning study is carried out to customise each of the RF models. The daytime, night-time and twilight precipitation events have to be treated separately due to differing information content about the cloud properties between the different times of day. Secondly, the RF models are trained using the optimum values for the number of trees and number of randomly chosen predictor variables found in the tuning study. Finally, the final RF models are used to predict rainfall rates using an independent validation data set and the results are validated against co-located rainfall rates observed by a ground radar network. To train and validate the model, the radar-based RADOLAN RW product from the German Weather Service (DWD) is used which provides area-wide gauge-adjusted hourly precipitation information.

Regarding the overall performance, as indicated by the coefficient of determination (Rsq), hourly rainfall rates show already a good correlation with Rsq = 0.5 (day and night) and Rsq = 0.48 (twilight) between the satellite and radar based observations. Higher temporal aggregation leads to better agreement. Rsq rises to 0.78 (day), 0.77 (night) and 0.75 (twilight) for 8-h interval. By comparing day, night and twilight performance it becomes evident that daytime precipitation is generally predicted best by the model. Twilight and night-time predictions are generally less accurate but only by a small margin. This may due to the smaller number of predictor variables during twilight and night-time conditions as well as less favourable radiative transfer conditions to obtain the cloud parameters during these periods.

However, the results show that with the newly developed method it is possible to assign rainfall rates with good accuracy even on an hourly basis. Furthermore, the rainfall rates can be assigned during day, night and twilight conditions which enables the estimation of rainfall rates 24 h day.

* Corresponding author at: Philipps-University Marburg, 35037 Marburg, Germany. Tel.: +49 6421 2825954; fax: +49 6421 2825670.
E-mail address: meike.kuehnlein@staff.uni-marburg.de (M. Kühnlein).

0034-4257
http://dx.doi.org/10.1016/j.rse.2013.10.026

1. Introduction

Many ecological and biodiversity-oriented projects require area-wide information on precipitation distribution and quantity at high temporal and spatial resolution. For this purpose satellite-based rainfall retrievals are often the only option. Traditionally, precipitation is observed locally by conventional instruments such as rain (or snow) gauges and where available, weather radar systems. However, the most obvious limitations of station-based precipitation measurements are their spatial incoherencies, uneven global distribution and highly variable density with some regions having a relatively dense network while others have only a few or no gauges. Over oceans, gauges are almost non-existent apart from a few island locations. Ground-based weather radar systems like the ones in Europe, Japan or North America provide spatial measurements of precipitation (100 km from the radar). They are located over land and generally concentrated in regions that are also well covered with rain gauges. In this context, precipitation retrievals from optical sensors aboard geostationary (GEO) weather satellites may be an alternative to fill the information gap by providing area-wide data about rainfall distribution and amount at high spatial and temporal resolutions.

During the last decades, several satellite-based rainfall retrieval techniques for the detection of precipitating clouds and assignment of rainfall rates from optical sensors available on GEO platforms have been developed (see valuable overviews by e.g. Kidd & Huffman, 2011; Kidd & Levizzani, 2011; Prigent, 2010; Thies & Bendix, 2011). Traditionally, GEO system-based retrieval schemes were restricted by the spectral resolution of the sensors, which only allowed concepts that rely on a relationship between cloud top temperatures measured in an infrared (IR) channel and rainfall probability/intensity (e.g. Adler & Mack, 1984; Arkin & Meisner, 1987). These concepts are based on the assumption that cold cloud tops are associated with high rainfall probabilities/intensities. More advanced IR retrieval methods are able to divide these precipitating cloud areas into different sub-areas, to which rainfall intensities are assigned (e.g. Adler & Negri, 1988; Hong, Hsu, Sorooshian, & Hiaogang, 2004; Hsu, Gao, & Sorooshian, 2002; O'Sullivan, Wash, Stewart, & Motell, 1990; Porcu & Levizzani, 1992; Reudenbach, 2003; Reudenbach, Nauss, & Bendix, 2007; Wu, Weinman, & Chin, 1985). Such IR retrievals show good results in regions with mainly convective clouds, especially in the tropics and sub-tropics, but exhibit considerable drawbacks concerning the detection and quantification of rainfall from stratiform clouds in connection with extra-tropical cyclones (e.g. Adler, Kidd, Petty, Morissey, & Goodman, 2001; Amorati, Alberoni, Levizzani, & Nanni, 2000; Ebert, Janowiak, & Kidd, 2007; Früh et al., 2007; Levizzani, Porcu, & Prodi, 1990; Negri & Adler, 1993; Pompei, Marrocu, Boi, & Dalu, 1995). This type of precipitating clouds is characterised by relatively warm and spatially homogeneous cloud top temperatures that do not differ significantly from raining to non-raining regions. Therefore, retrieval techniques based solely on IR cloud top temperature led to uncertainties concerning the assigned rainfall rate (e.g. Ebert et al., 2007; Früh et al., 2007).

With the upcoming of new generation GEO systems, several authors suggested the use of optical and microphysical cloud parameters derived from the now available multispectral data set to improve optical rainfall retrievals (e.g. Ba & Gruber, 2001; Kühnlein, Thies, Nauss, & Bendix, 2010; Nauss & Kokhanovsky, 2006; Roebeling & Holleman, 2009; Rosenfeld & Gutman, 1994; Rosenfeld & Lensky, 1998; Thies, Nauss, & Bendix, 2008a, 2008b, 2008c). They were able to show that cloud areas with a high optical thickness and a large effective particle radius possess a high amount of cloud water, and are characterised by a higher rainfall probability and intensity than cloud areas with a low optical thickness and a small effective particle radius.

Thies et al. (2008c) showed the possibility to separate areas of differing precipitation processes and rainfall intensities within the rain area by means of cloud properties retrieved with SEVIRI aboard MSG. The day and night technique for precipitation process separation and rainfall rate differentiation relies on information about the cloud top height, the cloud water path and the cloud phase in the upper parts. It is based on the assumption that areas with higher cloud water path and more ice particles in the upper parts are characterised by higher rainfall intensities. Recently, Kühnlein et al. (2010) used MSG SEVIRI reflection values in the 0.56–0.71 μm ($VIS_{0.6}$) and 1.5–1.78 μm ($NIR_{1.6}$) channels, which provide information about the optical thickness and the effective radius, to estimate rainfall rates over the northern German lowlands. This approach is based on the assumption that high rainfall rates are linked to high optical thickness and large effective particle radius, whereas low rainfall rates are linked to a low optical thickness and a small effective particle radius. The encouraging validation results of both retrievals indicate the high potential for an improved rainfall rate retrieval in the mid-latitudes using optical and microphysical cloud properties derived from MSG SEVIRI data which provide the enhanced spectral resolution that is needed (Levizzani, 2003; Levizzani et al., 2001). Furthermore, the relatively high spatial (3 km × 3 km at sub satellite point) and especially temporal resolution (15 min) permits a quasi-continuous observation of rainfall distribution and rainfall rate in near-real time.

In order to relate the retrieved cloud properties to precipitation, parametric statistics are generally used (e.g. Adler & Negri, 1988; Cheng & Brown, 1995; Kühnlein et al., 2010; Levizzani et al., 1990; Thies et al., 2008c). The application is rather straight-forward if only a few input variables are considered. However, cloud-top properties to precipitation relationship is very complex and non-linear, and it is likely beyond the skill of parametric tests and the related conceptual model.

Machine learning algorithms such as support vector machines (Mountrakis, Im, & Ogole, 2011), artificial neural networks (Mas & Flores, 2008), decision trees (Breiman, Friedman, Olshen, & Stone, 1984) or ensemble classifiers (Breiman, 1996) have been successfully adopted to remote sensing and rainfall applications (Capacci & Conway, 2005; Grimes, Coppola, Verdecchia, & Visconti, 2003; Rivolta, Marzano, Coppola, & Verdecchia, 2006) and may be suitable to overcome the limitations of the parametric techniques. When faced with high dimensional and complex data, machine learning algorithms provide efficient alternatives and generally show a higher accuracy (Foody, 1995; Friedl & Brodley, 1997; Hansen, Dubayah, & Defries, 1996). In addition, the developments in parallel computing with machine learning offer new possibilities in terms of training and predicting speed resulting in improved real time systems.

In recent years, machine learning techniques which use ensembles of classifications or regressions (e.g. random forests, neural network ensembles, bagging and boosting, see Friedl, Brodley, & Strahler, 1999; Krogh & Vedelsby, 1995; Rodriguez-Galiano, Ghimire, Rogan, Chica-Olmo, & Rigol-Sanchez, 2012; Ruiz-Gazen & Villa, 2007; Steele, 2000) have received increasing interest. They are based on the assumption that a whole set of trees or networks produce a more accurate prediction than a single tree or network (Dietterich, 2002). A new, powerful and promising ensemble classification and regression technique is random forests (Breiman, 2001). It is one of the most accurate learning algorithms available and it offers specific features that make it attractive for remote sensing applications. For example, it runs efficiently on large data sets, it is simple and can easily be applied to parallel computing platforms and it can capture non-linear association patterns between predictors and response. Although widely applied in other disciplines such as bioinformatics (e.g. Cutler & Stevens, 2006), some land-cover classifications using hyper spectral and multispectral satellite data, radar and lidar data (Ghimire, Rogan, & Miller, 2010; Guo, Chehata, Mallet, & Boukir, 2011; Pal, 2005; Rodriguez-Galiano et al., 2012) and also in a few ecological studies (e.g. Cutler et al., 2007; Mota, Jiménez, Amate, & Peñas, 2002; Prasad, Iverson, & Liaw, 2006), the utilisation of random forests in climatology remains rare. This is one of the reasons it has led us to investigate the usefulness of RF approaches for rain rate delineation from satellite platforms.

In summary, the enhanced information content on cloud properties at high spectral, spatial and temporal resolution offered by current and

upcoming GEO systems along with the encouraging results concerning rainfall rate differentiation and rainfall rate assignment shown by Thies et al. (2008c) and Kühnlein et al. (2010) point to a promising potential of optical sensors as basis for reliable rainfall rate retrievals. However, this potential will likely remain unexploited using common parametric approaches. In addition, and not explicitly stated above, all existing optical retrievals that are based on optical and microphysical cloud parameters are restricted to daytime and night-time conditions and do not cover twilight conditions (see e.g. Kidd & Levizzani, 2011; Prigent, 2010).

In this study we show the potential of the random forests method for an improved rainfall rate assignment during day, night and twilight based on MSG SEVIRI data which provide information on cloud properties at high temporal and spatial resolution.

Based on the precipitation processes in connection with extra-tropical cyclones, the retrieval process consists of three steps: (i) identification of precipitating cloud areas, (ii) separation of precipitating areas into predominantly convective and advective-stratiform cloud regions, (iii) individual assignment of rainfall rates to these cloud areas. Since this study focuses on an improved assignment of rainfall rates based on random forests and not on the development of an optimised precipitation retrieval (which includes the delineation of raining from non-raining areas), radar data is used for the first two steps. This means that the derivation of the rain area as well as the rain process is based on observations from the radar network rather than MSG SEVIRI. Rainfall rates are then assigned to the already identified stratiform and convective precipitating areas.

Germany was chosen as study area for the development and validation of the new technique. The region can be regarded as sufficiently representative for mid-latitudes precipitation formation processes since it is dominated by frontally induced precipitation processes in connection with extra-tropical cyclones and shows a prominent maritime to continental gradient from west to east. Moreover, the radar network based and gauge-adjusted, hourly precipitation data set (RADOLAN RW) provided by the German Weather Service (DWD) provides a reliable training and validation basis.

The structure of this paper is as follows: the underlying data sets and methods are introduced in Section 2. Section 3 gives a presentation of the theoretical background and conceptual design providing the basis for the selection of random forests predictors. In Section 4, the adjustment of the random forests models as well as the appraisal of the new rainfall rate assignment technique is introduced. The paper is closed with a summary and some conclusions in Sections 5.

2. Data and methods

2.1. Satellite observations

For this study, MSG SEVIRI data are used. SEVIRI scans the full disk every 15 min and measures reflected and emitted radiance in 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. The nominal spatial resolution at the sub-satellite point is 1 by 1 km for the high-resolution channel, and 3 by 3 km for the other channels (Aminou, 2002; Schmetz et al., 2002). Over the study area in Germany, the satellite viewing zenith angles of SEVIRI range from 56° to 64°. As a consequence, the above mentioned spatial resolution is reduced in the present study. The follow up mission Meteosat Third Generation is intended to be launched in 2018 (EUMETSAT, 2013). This ensures these data availability and utilisation of applications developed for MSG SEVIRI for the next decades. The MSG SEVIRI data required for this study were downloaded from the EUMETSAT data centre (www.eumetsat.int). Processing has been performed based on a newly designed Meteosat processing scheme which has been implemented by Tobias Ebert and Johannes Drönner in co-operation with the working group of Bernhard Seeger from the computer science department at Marburg University. The processing scheme will be available online at http://umweltinformatik-marburg.de/software/ shortly. Until then please contact the author for a copy of the software.

Cloud properties such as cloud effective radius and cloud optical thickness are retrieved using the semi-analytical approach SLALOM (SimpLe Approximations for cLOudy Media) developed by Nauss and Kokhanovsky (2011). This forward model is based on approximated solutions of the asymptotic radiative transfer theory (e.g. Germogenova, 1963; King, 1987) and provides increased computation speed since the equations can be efficiently solved during runtime. In order to retrieve cloud optical thickness and cloud effective droplet radius from MSG SEVIRI data, a combination of reflectance measurements at visible (0.65 μm) and near-infrared (1.64 μm) wavelengths is used. For the background albedo, a minimum composite of the reflectance in the visible (0.65 μm) and near-infrared (1.64 μm) channel over one month was calculated. A validation of SLALOM over sea and land surface against the well-known NASA MODIS cloud property product (Platnick et al., 2003) as well as the CloudSat 2B-TAU product (Polonsky, Labonnote, & Cooper, 2008) showed good agreement and can be found in Kühnlein, Appelhans, Thies, Kokhanovsky, and Nauss (2013). The present version of SLALOM is limited to water clouds. Cloud masks and cloud phase were derived using the algorithm by Cermak (2006) and Cermak and Bendix (2008) which have kindly been provided by the authors and are also implemented in the new Meteosat processing scheme.

2.2. Weather radar observations

For the development and validation of the new rainfall rate technique, radar-based precipitation data of the German Weather Service is used. The RADOLAN RW product is based on measurements with a C-band Doppler radar. Rain intensity adapted Z–R relationships, statistical clutter filtering and shadowing effects are treated within an online calibration process. Furthermore, precipitation intensities are adapted with ground-based precipitation measurements. The final precipitation product is available at temporal resolution of one hour and is a composite consisting of 16 German radar stations and some from neighbouring countries (e.g. Nancy/France) covering the entire area of Germany at a spatial resolution of 1 by 1 km (Bartels et al., 2004).

2.3. Pre-processing of satellite and weather radar observations

The different temporal and spatial characterises of the satellite data (15 min; 3 by 3 km at sub-satellite point) and weather radar data (more or less continuously over 1 h; 1 by 1 km) must be addressed to ensure the pixel matching between satellite and radar data. An average of the satellite-based products is aggregated over a time interval of one hour. This is done by taking the arithmetic mean of the four scenes available every hour. To assure that only cloudy pixels within the time interval are incorporated, the cloud mask developed and implemented by Cermak (2006) and Cermak and Bendix (2008) is applied. Only those SEVIRI pixels that are classified as cloudy over the entire time interval are taken into account. Because of the differing viewing geometries between both systems, the radar product was projected and spatially aggregated (mean) to the geometry of SEVIRI.

The final data set consists of 1150 scenes of precipitation events between April and September 2010. Scenes with at least 2000 rainy pixels were chosen as precipitating events based on the RADOLAN RW product. Hereby pixels with higher than 0.06 mm/h are considered as rainy. The data set is split into daytime, night-time and twilight data sets. To ensure sufficient solar illumination in the VIS and NIR channels, scenes with a corresponding solar zenith angle less than 70° belong to the daytime data set. 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-time data set. The resulting daytime

data set consists of 525 scenes, the night-time data set has 274 and the twilight data set has 351 scenes.

2.4. Random forests

The ensemble technique random forests which has been shown to perform very well in a variety of environmental investigations, contains a combination of characteristics that make it well suited for its application in remote sensing. RF runs efficiently on large data sets, can easily be parallelised and is relatively robust to outliers and noise. Furthermore, it does not require the specification of an underlying data model, it offers the ability to capture non-linear association patterns between predictors and response and it is able to deal with highly correlated predictor variables. It also generates an internal unbiased estimate of the generalisation error (OOB error) and has the ability to determine which variables are important in the regression. Finally, it offers the flexibility to perform several types of statistical data analysis (e.g. regression and classification) and it is computationally lighter than other tree ensemble methods.

Below, a brief overview of the RF procedure is given. For more theoretical details the reader is referred to the literature (Boulesteix, Janitza, Kruppa, & König, 2012; Breiman, 2001; Breiman & Cutler, 2008; Malley, Malley, & Pajevic, 2011; Strobl, Malley, & Tutz, 2009).

In general, the RF algorithm for regression works as follows:

1. *ntree* bootstrap samples are randomly selected from the data set with replacement. For each bootstrap, a different subset of the data set is used to develop the decision tree model. About one-third of the cases are left out of the sample. This out-of-bag (OOB) is used to get unbiased estimates of the regression error and to get estimates of the importance of the variables used for constructing the tree.
2. A regression tree for each of the bootstrap samples is grown (resulting in *ntree* trees) with the following modification: at each node, a subset of the predictor variables (*mtry*) is selected randomly to create the binary rule. In other words, *mtry* specifies the number of randomly chosen variables upon which the decision for the best split at each node is made. Variable selection is based on the residual sum of squares i.e. the predictor with the lowest residual sum of squares is chosen for the split. *mtry* is held constant during the forest growing.
3. Each of the *ntree* trees is grown to the largest extent possible. There is no pruning.
4. Finally, predictions are calculated by putting each OOB observation or observation of the test data set down each of the *ntree* trees. Then the predictions of all regression trees are averaged to produce the final estimate (Breiman, 2001).

The OOB error is an important feature of RF. As mentioned before, each tree is built on a bootstrap sample that comprises roughly two-third of the training data. The remaining one-third (OOB) of the training data is not included in the learning sample for this tree and can be used to test it. Therefore, the RF model is applied to the OOB data. Then, the deviations between predicted and observed values are used to calculate the OOB error, which is for regression the mean square error (MSE), and is given by

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(RR_{Predi} - RR_{Obsi})^2 \quad (1)$$

where RR_{Predi} is the *i*th prediction and RR_{Obsi} is the *i*th observation. This resulting OOB error provides an unbiased estimate of the generalised error and can be considered as an internal validation. As long as enough trees have been grown, OOB's estimate of error rate is quite accurate (Breiman, 1996, 2001). Otherwise the OOB estimate can bias upward (see Bylander, 2002).

While bagging uses all predictor variables at each node (bagging = mtry = number of predictors), RF constructs a tree using different training data subsets created through bagging and bootstrap of the data. By making the tree grow from different bootstrap samples, the diversity of the trees is increased. This increases generality, makes the regression more robust when facing slight variations in the training data and generally increases the overall prediction accuracy (Breiman, 2001). Several studies have shown that methods based on bagging are not sensitive to noise (Briem, 2002; Chan & Paelinckx, 2008). When RF makes a tree grow, it uses the best split based on a number of randomly sampled predictor variables. If all variables were used for each tree, the trees would be very identical and therefore highly correlated (Breiman, 2001). Thus, the randomly chosen subsets of predictor variables at each split of each tree ensure lower correlation between trees that in turn increases model robustness.

Beside the favourable features of RF, some limitations need also be mentioned. One of the most significant drawbacks of RF is that it does lack interpretability. Since the predictions are derived using a forest of trees, it is not possible to easily illustrate how the predictions are made (i.e. no single tree can be drawn to illustrate the decisions upon which the predictions are based). Furthermore, averaging over all trees means that it is neither possible to predict beyond the range of response values in the training data, nor to predict the entire range of response values. As a result, RF tends to overestimate low values and underestimate high values. Furthermore, RF is a truly random statistical method which entails a number of methodological issues related to repeatability and generalisability of the analyses. For in depth descriptions and generally accepted solutions of how to address these issues, the reader is referred to the fundamental statistical literature provided for the method. At the very least, the complete RF procedure needs to be repeated several times to evaluate the general robustness of the obtained predictions.

For our calculations we used an R implementation of the RF library, which was created by Liaw and Wiener (2002) based on the original Fortran code by Leo Breiman and Adele Cutler (for information on the open source software R see R Development CoreTeam (2008)). The package is called "randomForest". The algorithm falls into the *embarrassingly parallel* category. This means that the number of trees to grow within a RF model can be divided into independent subsets, since each tree in the forest depends only on the given data set and not on the other trees. The subsets can be built on all available cores or on different machines. Then the resulting RF objects are combined to get the final forest. The parallel execution was realised by using the "foreach" and "doSNOW" packages. The R code used for the analysis at hand is available on our homepage (http://umweltinformatik-marburg.de/software/).

3. Selection of predictor variables

Similar to parametric approaches, RF also requires a set of predictor variables to estimate a response variable. While the selection of the response variable i.e. the rainfall rate is obvious for the study objective, the definition of predictor variables should reflect the conceptual framework of the rainfall assignment technique which in turn must consider the mid-latitudinal precipitation processes with a strong focus on extra-tropical cyclones.

Houze (1993) summarised the conceptual model of rainbands dominated by different rainfall processes in connection with extra-tropical cyclones. Following this conceptual model of rainbands, the precipitation field can be decomposed into areas dominated by different rainfall processes: (i) advective-stratiform background and intermediary precipitation which are linked to light precipitation intensities (further referred as advective-stratiform precipitation process) and (ii) narrow cold-frontal, wide cold-frontal and warm-frontal rain bands which are characterised by high rainfall intensities (further referred as convective precipitation process). Hence, rainbands dominated by different rainfall

processes lead to differing rainfall rates. For this reason the rainfall rate assignment is realised in a three step approach (identification of precipitating cloud areas, separation according to their process, final rainfall rate assignment).

Regarding the vertical cloud extension, convectively dominated precipitation areas with higher rainfall intensities are characterised by larger cloud depths and cloud tops reaching higher into the troposphere. On the other hand, advective-stratiform precipitation areas are not necessarily connected to cold cloud top temperatures and are also not colder than surrounding non-precipitating cloud areas which explains the limited accuracy of the traditional IR retrievals presented in the introduction. Therefore, several authors have successfully used either optical and microphysical cloud parameters derived from multispectral satellite data or a suitable selection of spectral channels and channel combinations which provide information about cloud parameters, for an improved rain area delineation (Lensky & Rosenfeld, 2003; Nauss & Kokhanovsky, 2006, 2007; Rosenfeld & Lensky, 1998; Thies et al., 2008a, 2008b), rainfall intensity differentiation (Thies et al., 2008c) and rainfall rate assignment (Kühnlein et al., 2010). They were able to show that cloud areas with a high cloud water path (i.e. large enough combination of the optical thickness and effective particle radius) possess a higher rainfall probability and higher rainfall rates than cloud areas with a small cloud water path. Since effective rain formation processes are mainly coupled to ice particles in the upper parts of the clouds and the "seeder-feeder" effect (Houze, 1993), advective-stratiform precipitation areas with a higher cloud water path and a higher amount of ice particles in the upper cloud regions are also characterised by higher rainfall intensities (Thies et al., 2008c).

Considering the dominant precipitation processes for convective and advective-stratiform rainfall areas within extra-tropical cyclones, the following cloud physical parameters are chosen for this study:

- Cloud top height (CTH)
- Cloud top temperature (CTT)
- Cloud phase (CP)
- Cloud water path (CWP)

A proper SEVIRI spectral channel selection can be used as surrogates for these cloud physical parameters and therefore as predictor variables for the random forest model.

As a good proxy for the cloud top temperature, the brightness temperature in the 10.8 μm channel ($BT_{10.8}$) can be used. In addition to the cloud top temperature, the brightness temperature difference between the water vapour (WV) and the IR channels are used to gain information about the cloud top height relative to the tropopause level which enables a reliable identification of deep convective clouds (Heinemann, Reudenbach, Heuel, Bendix, & Winiger, 2001; Schmetz, Tjemkes, Gube, & van de Berg, 1997; Tjemkes, van de Berg, & Schmetz, 1997). Thies, Nauss, and Bendix (2008d) showed that the WV–IR combinations of MSG SEVIRI perform differently for different cloud-top height to tropopause level settings. To include different sensitivities on cloud top height the two channel differences $\Delta T_{WV6.2-IR10.8}$ and $\Delta T_{WV7.3-IR12.1}$ have been chosen.

Since effective rain formation processes are mainly coupled to ice particles in the upper parts of the cloud, the cloud phase is incorporated. The channel differences between 8.7 μm and 10.8 μm ($\Delta T_{8.7-10.8}$) as well as between 10.8 μm and 12.1 μm ($\Delta T_{10.8-12.1}$) can be used to gain information about the cloud phase (Ackermann et al., 1998; Strabala, Ackerman, & Menzel, 1994; Thies et al., 2008a). At these two wavelengths, the absorption of ice and water is different (Baum & Platnick, 2006). The increase of water particle absorption is greater between 11 and 12 μm than between 8 and 11 μm. On the other hand the increase of ice particle absorption is greater between 8 and 11 μm than between 11 and 12 μm (Strabala et al., 1994). Therefore the difference $\Delta T_{8.7-10.8}$ of ice clouds are greater than coincident $\Delta T_{10.8-12.1}$ differences and the opposite is true for water clouds.

For day-time observations, the SLALOM retrieval is used to derive optical and microphysical cloud properties such as cloud effective radius (a_{ef}) and cloud optical thickness (τ) which are used as predictor variables for advective-stratiform rainfall rates. Because the application of the commonly available cloud property retrievals to ice clouds requires the a-priori definition of ice particle geometries, which in turn heavily influences the retrieved values (see also Kokhanovsky & Nauss, 2005), the authors decided to use the reflectance at 0.6 μm ($VIS_{0.6}$), 0.8 μm ($VIS_{0.8}$) and 1.6 μm ($NIR_{1.6}$) channel directly as input variables for the assignment of rainfall rates to convectively dominated precipitation areas. For night-time, there are MSG SEVIRI retrievals that can compute optical and microphysical cloud properties (e.g. VISST–SIST algorithms developed by NASA Langley Cloud and Radiation Group). Unfortunately, these algorithms derive reliable properties only for non-raining optically thin clouds (Minnis et al., 2011). However, several case studies have shown that implicit information about microphysical and optical cloud properties is available in the emissive channels (Baum et al., 2000; Inoue, 1985; Lensky & Rosenfeld, 2003; Ou, Liou, Gooch, & Takano, 1993; Ou et al., 2002; Stone, Stephens, Platt, & Banks, 1990; Strabala et al., 1994). Based on these studies Thies et al. (2008d) demonstrated that the SEVIRI channel differences $\Delta T_{3.9-10.8}$, $\Delta T_{3.9-7.3}$, $\Delta T_{8.7-10.8}$ and $\Delta T_{10.8-12.1}$ provide information about the CWP that can be used for rainfall retrieval. Therefore, the respective channel differences are incorporated instead of retrieved cloud properties during night-time. During twilight conditions neither the reflectances of VIS and NIR channels (due to insufficient solar illumination) nor the SEVIRI channel differences $\Delta T_{3.9-10.8}$ and $\Delta T_{3.9-7.3}$ can be used as surrogates for the CWP. The 3.9 μm channel radiance contains both reflected solar radiance and thermal emitted radiance. To use the 3.9 μm channel the solar component must be quantified and eliminated. This itself has been investigated by several studies (Rao, Ou, & Liou, 1995; Rosenfeld & Lensky, 1998) and they showed that simplifications and assumptions are necessary which are only acceptable for a certain kind of cloud. To prevent misinterpretation, predictors containing the 3.9 μm channel are not used to gain information about the CWP and therefore there are no predictor variables available representing the CWP during twilight.

In summary, a different set of predictor variables with respect to the degree of solar illumination (day, night, twilight) and the dominant precipitation processes within each cloud region is utilised. An overview is given in Table 1. Since the RF approach is not limited to a certain number of predictor variables, all channels which are part of the aforementioned channel combinations (e.g. $\Delta T_{10.8-12.1}$), are also incorporated in the test.

Table 1
Overview of RF predictor variables.

	Daytime		Night-time		Twilight	
	Stratiform	Convective	Stratiform	Convective	Stratiform	Convective
CTH	$\Delta T_{6.2-10.8}$ $\Delta T_{7.3-12.1}$	$\Delta T_{6.2-10.8}$ $\Delta T_{7.3-12.1}$	$\Delta T_{6.2-10.8}$ $\Delta T_{7.3-12.1}$	$\Delta T_{6.2-10.8}$ $\Delta T_{7.3-12.1}$	$\Delta T_{6.2-10.8}$ $\Delta T_{7.3-12.1}$	$\Delta T_{6.2-10.8}$ $\Delta T_{7.3-12.1}$
CTH/CTT	$IR_{10.8}$	$IR_{10.8}$	$IR_{10.8}$	$IR_{10.8}$	$IR_{10.8}$	$IR_{10.8}$
CP	$\Delta T_{8.7-10.8}$ $\Delta T_{10.8-12.1}$	$\Delta T_{8.7-10.8}$ $\Delta T_{10.8-12.1}$	$\Delta T_{8.7-10.8}$ $\Delta T_{10.8-12.1}$	$\Delta T_{8.7-10.8}$ $\Delta T_{10.8-12.1}$	$\Delta T_{8.7-10.8}$ $\Delta T_{10.8-12.1}$	$\Delta T_{8.7-10.8}$ $\Delta T_{10.8-12.1}$
CWP	$VIS_{0.6}$ $VIS_{0.8}$ $NIR_{1.6}$ a_{ef} τ CWP	$VIS_{0.6}$ $VIS_{0.8}$ $NIR_{1.6}$	$\Delta T_{3.9-10.8}$ $\Delta T_{3.9-7.3}$	$\Delta T_{3.9-10.8}$ $\Delta T_{3.9-7.3}$		
SEVIRI channels	$WV_{6.2}$ $WV_{7.3}$ $IR_{8.7}$ $IR_{12.1}$	$WV_{6.2}$ $WV_{7.3}$ $IR_{8.7}$ $IR_{12.1}$	$WV_{3.9}$ $WV_{6.2}$ $WV_{7.3}$ $IR_{8.7}$ $IR_{12.1}$	$WV_{3.9}$ $WV_{6.2}$ $WV_{7.3}$ $IR_{8.7}$ $IR_{12.1}$	$WV_{6.2}$ $WV_{7.3}$ $IR_{8.7}$ $IR_{12.1}$	$WV_{6.2}$ $WV_{7.3}$ $IR_{8.7}$ $IR_{12.1}$

Abbreviations are as follows: CTH, cloud top height; CTT, cloud top temperature; CP, cloud phase; CWP, cloud water path.

4. Estimation of rainfall rates

4.1. General methodology

On the basis of the theoretical background introduced in Section 3, the rainfall rate assignment technique is realised in three steps: (i) precipitating cloud areas are identified, (ii) the precipitating cloud areas are separated into convective and advective-stratiform dominated precipitation areas and (iii) rainfall rates are assigned to the convective and advective-stratiform dominated precipitation areas, respectively.

As mentioned in the introduction, the focus of the present study lies upon the development of a new technique for the assignment of rainfall rates. In order to develop and evaluate such a rainfall rate assignment technique, the RADOLAN RW product of the DWD is used to realise the first and second step. This means that both the rain areas, as well as the rain processes, are derived from the radar observations rather than from satellite images. Precipitation areas with more than 1.8 mm/h are considered as convectively dominated and areas between 0.06 and 1.8 mm/h are considered as advective-stratiform dominated (see Thies et al., 2008c). Hence, the proposed technique aims to assign rainfall rates to precipitating cloud areas that are already identified as being convective and advective-stratiform.

For the development and validation of the technique, a data set consisting of 525 daytime, 274 night-time and 351 twilight satellite scenes of precipitation events between April and September 2010 are used. The data set is split into daytime, night-time and twilight data sets which have to be treated separately due to differing illumination conditions. During night-time the channels at visible and very near infrared wavelengths (0.6 to 1.6 μm) are not available. During twilight and daytime the use of the 3.9 μm channel is complicated due to the varying solar component in this channel. This means that depending on the time of the day different random forest models are built and adapted. Each of these three data sets are randomly split into training (¼ of the scenes) and validation data sets (¾ of the scenes). The training data sets are used to train the according RF model, whereas the validation data sets are used to validate the predictions afterwards. The precipitation events taken for training are independent from those taken for validation. Since the technique aims to assign rainfall rates to precipitating cloud areas already classified as convective or advective-stratiform, each data set is split into cases of convectively dominated rainfall and cases of advective-stratiform dominated rainfall and treated separately (day = day-C and day-S, night = night-C and night-S, twilight = twilight-C and twilight-S; where C = convective and S = advective-statiform).

The development of the rainfall rate technique is realised in three steps:

(1) Tuning: the optimal values of *ntree* and *mtry* are assessed for each RF model.
(2) Training: the RF models are trained using the optimum values of model parameters found in the tuning study.
(3) Validation: the RF models are applied to the validation data sets and the predicted rainfall rates are validated against co-located rainfall rates observed by the radar.

All these steps are described in the following sections.

4.2. Model tuning

There are basically two parameters to adjust in the R-package "randomForest", the overall number of trees in a forest (*ntree*) and the number of predictor variables randomly sampled for use at each split (*mtry*). The tuning is based on performance of OOB data.

An important consideration is how many trees to grow within the random forest model. Breiman (2001) suggested that the generalisation error converges as the number of trees increases. Adding more and more trees to the model does not result in over-adjustment. The main limitation of increasing *ntree* is the extra computation time. Therefore, the number of trees is not a real parameter in the sense that there is an optimum value, rather the number should be as large as computationally feasible.

To assess the optimal value of *ntree*, a large number of RF models using randomly selected subsets of each data set (day-C, day-S, night-C, night-S, twilight-C, twilight-S) is created. Each RF model is created using 1000 trees for all possible values of *mtry* (mtry = 1 to maximum). Then the MSE values of every possible value of *mtry* is averaged. Fig. 1 shows how the error rate changes with the number of trees for *mtry* equal to the minimum (*mtry* = 1), mean (*mtry* = 6) and maximum value of *mtry* (*mtry* = 12/15, for details see below). From around 500 trees onwards, the MSE of each data set stabilises and the addition of trees neither increases nor decreases the MSE. Therefore, the number of trees in the forest can be regarded as sufficient using *ntree* larger than 500. In order to reduce extra computation time, *ntree* is set to

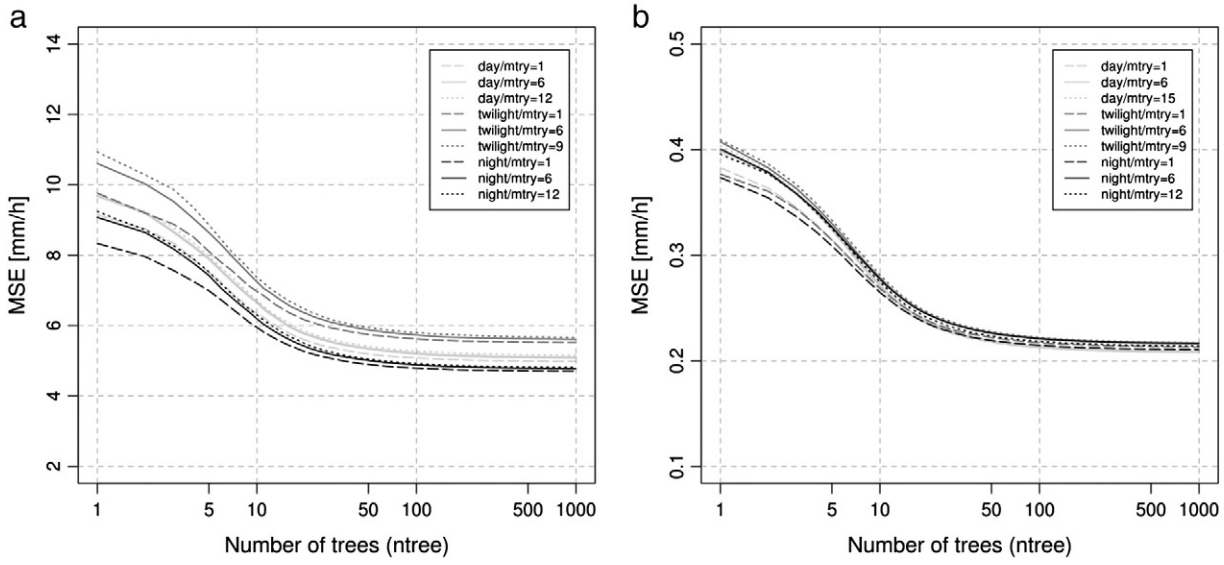

Fig. 1. Effect of number of trees (*ntree*) and random split variables (*mtry*) on OOB error (MSE). Results shown for (a) convective and (b) stratiform day, night and twilight tuning data sets, respectively.

500 in all models as this value is large enough to produce a stable prediction.

In comparison to *ntree*, the number of predictor variables randomly sampled for use at each split (*mtry*) is a real parameter. The suggested default value of *mtry* for regression is p/3 (with p total number of predictors) (Breiman, 2001). In practice the best values for *mtry* will depend on the problem at hand and the parameter should be treated as tuning parameter. If the recommended *mtry* is too small in the presence of a large number of noise predictors, then it is more likely to select non-informative predictors. On the other hand, a small value of *mtry* might offer the possibility for strong predictors to be chosen in a scenario with many informative predictors (Boulesteix et al., 2012). Reducing the number of predictor variables reduces the correlation between the trees, but also causes each individual tree to be more biased (Goldstein, Polley, & Briggs, 2011).

To assess the optimal values of *mtry*, a large number of RF models are created for each data set using different possible values of splitting variables (*mtry* = 1 to maximum number of *mtry*) while keeping *ntree* = 500 constant. First, the RF model is computed with *mtry* = 1. Then, *mtry* is increased by 1 and a new forest is built with the new *mtry*. This is done until the maximum number of *mtry* is reached. This whole process is repeated several times using randomly selected subsets of the data sets. The effect of *mtry* on the OOB error is exemplarily shown for the day-time models in Fig. 2. Based on these results, the value of *mtry* leading to the smallest OOB error of a forest is selected for the according RF model. Regarding the convective data sets, *mtry* = 7 leads to the highest prediction accuracy whereas *mtry* = 8 leads to the highest prediction accuracy for the stratiform data sets.

4.3. Model training

The training of the RF models is done on a pixel and hourly basis by a comparison with the ground-based radar precipitation data using the different training data sets. Hence, various satellite-based products (see Section 3) are combined with spatially and temporally co-located radar data from which the corresponding rainfall rates are extracted. The RF models are trained with *ntree* and *mtry* found in the tuning study (Section 4.2). All this is realised for daytime, night-time and twilight scenes as well as for convective and stratiform cases separately.

4.4. Model validation

Once the RF models are established, it is possible to assign rainfall rates based on the same predictors used for training. In order to assess the quality of the model, the established RF models are applied to the day, night and twilight validation data sets, respectively. Then the predicted rainfall rates (RR_{Pred}) are validated against the co-located rainfall rates observed by the radar (RR_{Obs}). For this, the hourly rainfall rates of RR_{Obs} and RR_{Pred} are also summed over 3-h and 8-h intervals where the latter contains all scenes of the day from the daytime, night-time and twilight time intervals respectively. For the appraisal of the retrieval technique, standard continuous verification scores are calculated. As measures of the agreement between observed and predicted values mean error (ME = bias), mean absolute error (MAE), root mean square error (RMSE) and coefficient of determination (Rsq) are calculated. These scores are given by:

$$ME = \frac{1}{N}\sum_{i=1}^{N}(RR_{Predi} - RR_{Obsi}) \tag{2}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|RR_{Predi} - RR_{Obsi}| \tag{3}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(RR_{Predi} - RR_{Obsi}\right)^2} \tag{4}$$

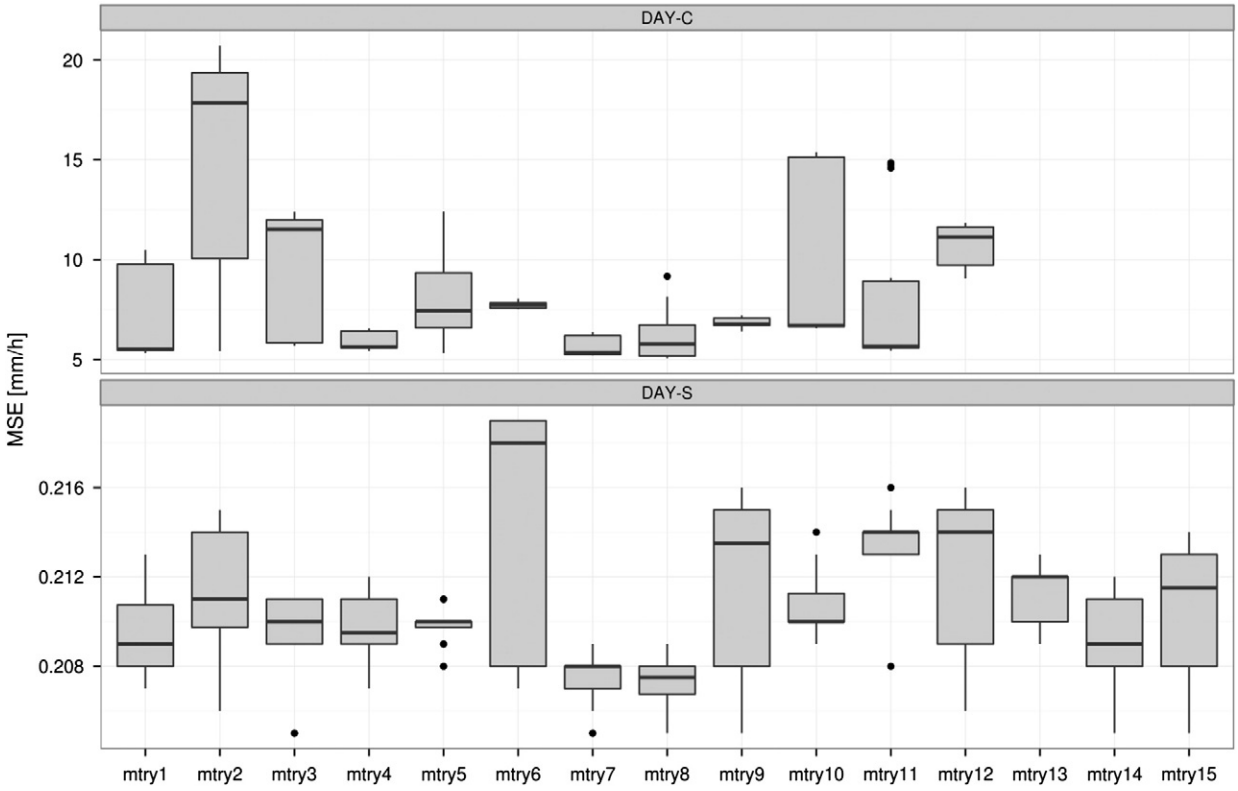

Fig. 2. Effect of the number of random split variables (*mtry*) on OOB error (using *ntree* = 500) for the daytime scenes (where C = convective and S = advective-statiform). Boxes indicate 25th, 50th and 75th percentiles. Whiskers extend to the most extreme data point within 1.5 times the interquartile range (75th-25th percentiles). Outliers shown as stars.

$$\mathrm{Rsq} = \left(\frac{\sum_{i=1}^{N} (RR_{Predi} - RR_{Pred})(RR_{Obsi} - RR_{Obs})}{\sqrt{\sum_{i=1}^{N} (RR_{Predi} - RR_{Pred})^2} \sqrt{\sum_{i=1}^{N} (RR_{Obsi} - RR_{Obs})^2}} \right)^2 \quad (5)$$

where RR_{Predi} is the ith prediction and RR_{Obsi} is the ith observation.

In order to assess the model performances four different evaluation strategies are employed:

(a) The overall performance of the rainfall rate assignment technique during daytime, night-time and twilight conditions is investigated.
(b) The performance of the rainfall rate assignment technique is investigated on a scene-by-scene basis.
(c) The diurnal performance of the RF models is investigated.
(d) The influence of different training data sets on the model performances is investigated.

First, the overall performance of the rainfall rate retrieval during daytime, night-time and twilight conditions is investigated. By extracting the data pairs of the validation data sets on a pixel basis, a total of 1309721 pairs of RR_{Obs} and RR_{Pred} are made available on an hourly basis, containing 678600 daytime, 280776 night-time and 350345 twilight data pairs. The hourly rainfall rates have also been aggregated for 3- and 8-h intervals. With increasing temporal aggregation, the number of data pairs decreases accordingly. The performance of the models across different aggregation times is summarised in Table 2. As Table 2 shows, the rainfall rates observed by the radar at a 1-h interval range between 0.06 and 56.10. However, the spread of rainfall rates predicted by RF is much smaller and ranges between 0.16 and 26.51 for the same time interval. This highlights that very high and low observations cannot be captured by RF as the response variable is calculated by putting each object down the decision trees and then averaging the predictions of all trees. As a consequence, the averaging reduces the very high rainfall rates and increase the very small rainfall rates, which leads to an underestimation of the high rainfall rates and overestimation of the small rainfall rates. A consistent under-estimation (over-estimation) of the maximum (minimum) is apparent for all sets of models, regardless of aggregation or time of day, especially maximum rainfall which can be severely under-estimated. However, mean, median and standard deviations generally show good agreement (please note, that it is also not possible to predict beyond the range of the response values in the training data).

As indicated by the Rsq, hourly RR_{Obs} and RR_{Pred} already show a good correlation with Rsq = 0.5 (day and night) and Rsq = 0.48 (twilight). Higher temporal aggregation leads to better agreement (Rsq as high as 0.78 (day), 0.77 (night) and 0.75 (twilight) for 8-h scenes). However, this is paralleled by an increased spread in the error scores. Small positive ME are seen for all daytime and twilight time intervals (ranging from 0.10 to 0.30), indicating a slight overestimation of the rainfall rate. During night-time, a slight under-estimation of rainfall rates are indicated (ME ranging from −0.07 to −0.13 for the different aggregation intervals).

In the next step the performance of the rainfall rate assignment technique on a scene-by-scene basis is investigated with results are summarised in Fig. 3. The standard verification scores calculated from RR_{Obs} and RR_{Pred} on a scene basis are plotted as box plots to facilitate a visual analysis of the performance.

There is a general pattern of daytime precipitation being best estimated by the model, regardless of aggregation times. Twilight and night-time predictions are generally less accurate but only by a small margin. Overall, there are no significant differences in central tendency arising from the time of day and this is remarkable particularly for

Table 2
Statistical values from the validation data sets.

		Predicted	Observed	Predicted	Observed	Predicted	Observed
Time period (tp)		1-h	1-h	3-h	3-h	8-h	8-h
Day	Min [mm/tp]	0.16	0.06	0.18	0.06	0.18	0.07
	Max [mm/tp]	26.51	56.10	32.91	49.43	48.05	80.00
	Median [mm/tp]	0.78	1.00	1.01	1.26	1.70	1.88
	Mean [mm/tp]	1.61	1.48	2.38	2.18	2.58	3.28
	Std [mm/tp]	1.57	1.68	2.54	2.67	4.10	4.28
	Rsq	0.50		0.68		0.78	
	RMSE [mm/tp]	1.26		1.58		2.07	
	MAE [mm/tp]	0.72		0.88		1.13	
	ME [mm/tp]	0.13		0.20		0.30	
	N	678600		422205		305871	
Night	Min [mm/tp]	0.14	0.06	0.14	0.06	0.14	0.06
	Max [mm/tp]	9.02	40.13	18.53	49.00	27.55	62.12
	Median [mm/tp]	0.88	1.09	1.66	1.79	1.86	1.96
	Mean [mm/tp]	1.76	1.83	2.87	3.01	3.38	3.51
	Std [mm/tp]	1.43	2.07	2.65	3.55	3.39	4.34
	Rsq	0.50		0.69		0.77	
	RMSE [mm/tp]	1.46		2.00		2.13	
	MAE [mm/tp]	0.80		1.06		1.13	
	ME [mm/tp]	−0.07		−0.14		−0.13	
	N	280776		149608		146136	
Twilight	Min [mm/tp]	0.18	0.06	0.22	0.07	0.19	0.07
	Max [mm/tp]	21.02	47.58	26.03	52.62	32.79	58.24
	Median [mm/tp]	0.87	1.085	1.54	1.65	1.66	1.76
	Mean [mm/tp]	1.79	1.69	2.87	2.71	3.30	3.12
	Std [mm/tp]	1.58	1.80	2.83	3.09	3.49	3.73
	Rsq	0.48		0.68		0.75	
	RMSE [mm/tp]	1.35		1.78		1.91	
	MAE [mm/tp]	0.79		1.03		1.10	
	ME [mm/tp]	0.10		0.16		0.18	
	N	350345		190758		145857	

The scores are based on data pairs of 394 daytime, 205 nighttime and 264 twilight precipitation scenes from April to September 2010. The values are calculated from the RR_{Obs} and RR_{Sat} data pairs extracted from the whole day, night and twilight data set, respectively. "Min" and "Max" denote the minimum and maximum value. "Mean" and "Median" signify the average and median value. "Std" signifies the standard deviation. Abbreviations are as follows: Rsq, coefficient of determination; RMSE, root mean square error; MAE, mean absolute error; ME, mean error; N, number of data pairs considered.

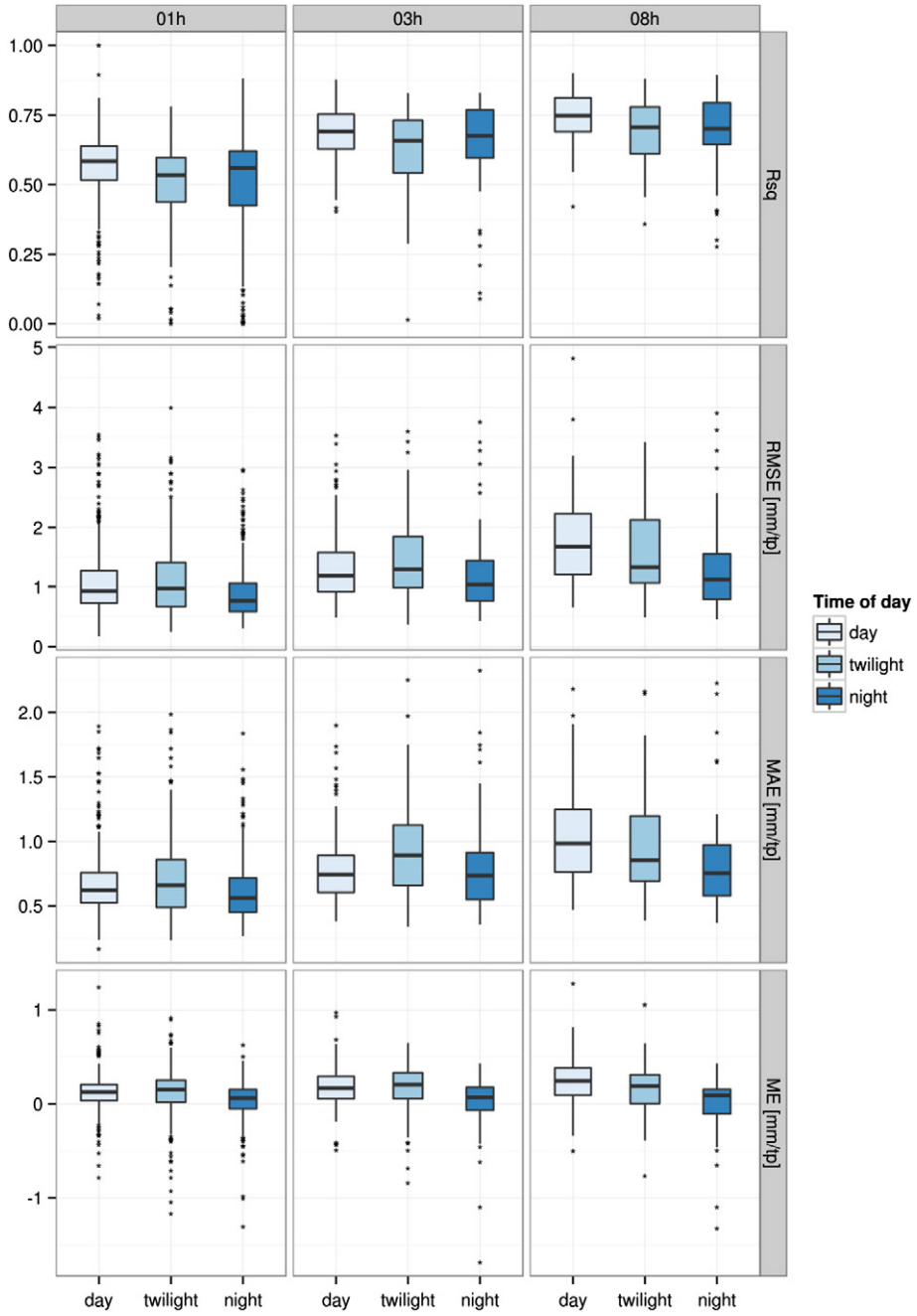

Fig. 3. Performance of rainfall rate assignment on a scene by scene basis. Box and whisker plots showing distribution of standard verification scores (rows) of RR_{obs} vs. RR_{pred} for different aggregation times (columns) according to time of day (colours). Boxes indicate 25th, 50th and 75th percentiles. Whiskers extend to the most extreme data point within 1.5 times the interquartile range (75th–25th percentiles). Outliers shown as stars. Abbreviations are as follows: Rsq, coefficient of determination; RMSE, root mean square error; MAE, mean absolute error; ME, mean error; tp, time period.

twilight conditions. The most obvious difference between different times of day lies in the consistency of the prediction performance indicated by the spread of the Rsq. Here, twilight and night-time scenes clearly show greater variability, as indicated by the elevated inter quartile ranges (i.e. height of the boxes). As mentioned earlier, elevated aggregation exhibits better agreement between predictions and observations, which is paralleled by an increase in the error scores in both magnitude and range. A look at the ME distribution reveals a general tendency of the model to overestimate precipitation, especially during day and twilight hours. Night-time predictions are more balanced in this respect. In general, it becomes apparent that night-time errors are smaller than during other times of the day.

Figs. 4a and 5a show examples of hourly rainfall rates observed by the radar and predicted by RF model. Both figures illustrate that for higher RR_{Obs} (>10 mm/h), the corresponding RR_{Pred} are predominantly smaller. At the same time, the areas in the surrounding of high rainfall areas are slightly overestimated by RF. As a result of the aggregation process, it is possible to better reproduce the observed rainfall rate (Figs. 4b and 5b) but again, very high rainfall rates (>20 mm/3 h) cannot be captured by the model. As already mentioned, the observed mismatch between predicted and observed extremes is characteristic of random forests.

General diurnal performance of the RF model is shown in Fig. 6. Here, all model verification scores are shown for each hour of the day

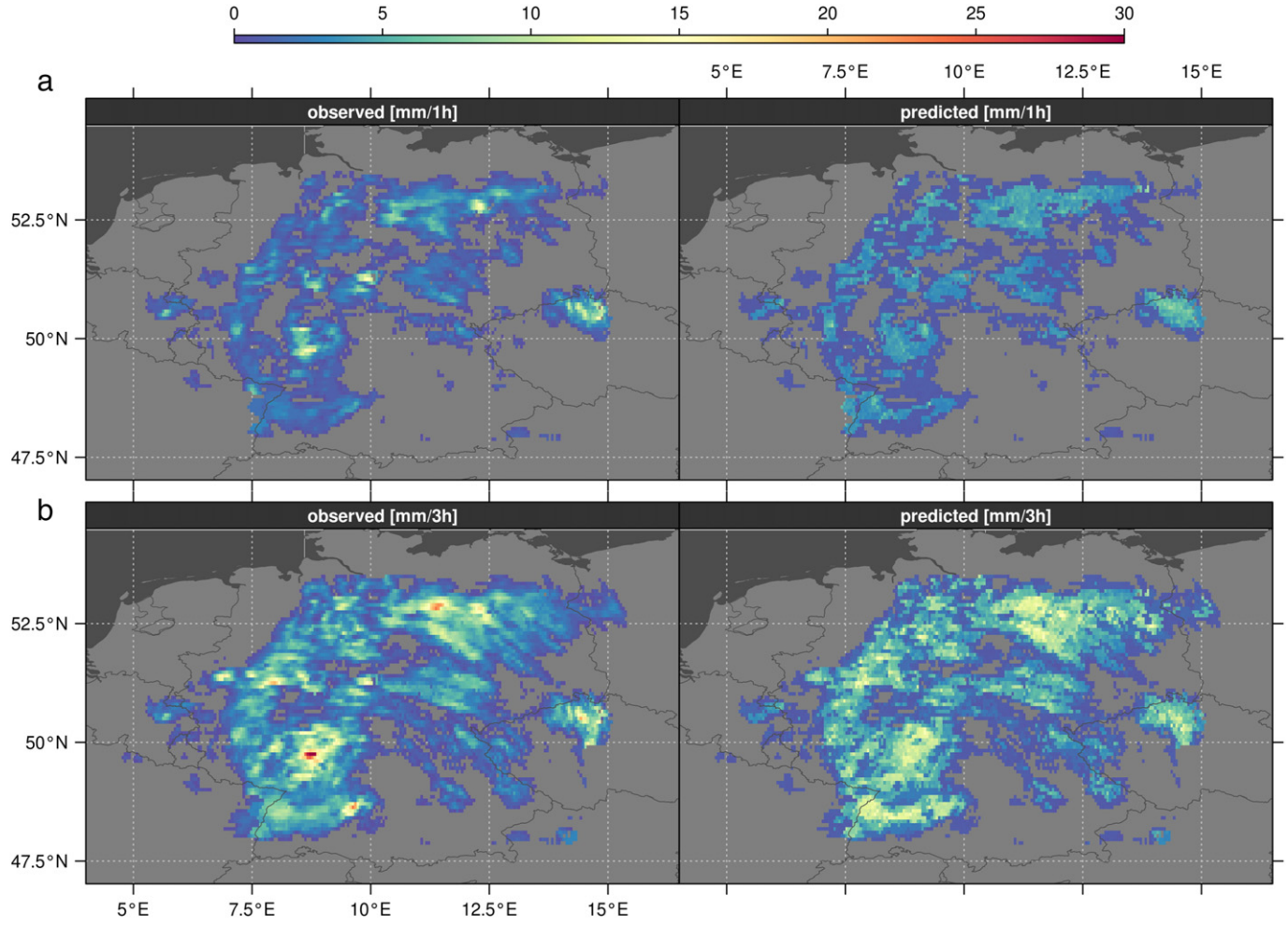

Fig. 4. Rainfall rates observed by radar (left) and predicted by RF model (right) during daytime. (a) rainfall rates for scene from 6 May 2010 15:00 UTC, (b) aggregated rainfall rates for scenes from 6 May 2010 14:00 to 16:00 UTC.

across all scenes. The comparison reveals reasonable agreement between RR_{Obs} and RR_{Pred} for all hours of the day. Best performance is seen during late morning and midday (increased Rsq and reduced spread in the error measures) most likely because of the favourable solar observation geometries. Twilight hours reveal rather variable prediction performance, especially toward the end of the day, whereas night-time performance is rather consistent. The maximum divergences occur during twilight conditions. It is to be expected that performance is different. Depending on solar illumination in twilight conditions the retrieval of cloud properties is more difficult and there are fewer predictor variables that can be used due to solar illumination. Furthermore, there are a reduced number of scenes which is indicated by the width of the boxes.

As already mentioned, the data used for the training of the model may influence the prediction. In order to assess how different input data sets may influence the RF results we have repeated the complete analysis 10 times, each with different randomly split sets of independent training and validation sets. For each of these iterations, the same standard verification scores as above are calculated on a pixel-basis and presented in Fig. 7 according to the time of the day. The daytime and night-time validation data sets show the highest median values for all aggregation levels with the night-time validation data set having a bigger range of Rsq values with increased temporal aggregation. By comparison to day and night, the twilight results are slightly below. The Rsq values are less and the range of Rsq values is bigger. As can be seen in column 2 and 3, the correlations increase considerably at lower temporal resolutions of 3- and 8-h. These results indicate that the selection of randomly chosen training and validation data sets influences the overall prediction accuracy, but in a small margin. This could be expected since the observations of the training sets are used to build the RF model and therefore influence the prediction. However, even if the results show that the different training sets influence the overall prediction, the standard verification scores reveal the same patterns which were shown in the foregoing investigations. This means that they reflect the same behaviour according to temporal aggregation and time of day.

5. Summary and conclusions

The aim of the present study was to investigate the potential of MSG SEVIRI for improved rainfall rate assignment using the ensemble classification and regression technique random forests as a fundamental algorithm. The novel approach differs from the most state-of-the-art satellite-based rainfall retrievals since it is not using a conventional parametric approach but a machine learning algorithm. RF is one of most accurate learning algorithm available and offers specific features that make it attractive for remote sensing applications, e.g. it runs efficiently on large data sets, it is simple and easy to parallelise. One of the key advantages is the ability to capture non-linear association patterns between predictors and response, which becomes important when dealing with a very complex non-linear event like precipitation. Due to the deficiencies of existing rainfall retrieval techniques based on the IR cloud-top temperature concerning the detection and quantification of rainfall from stratiform clouds, the aim of the present study

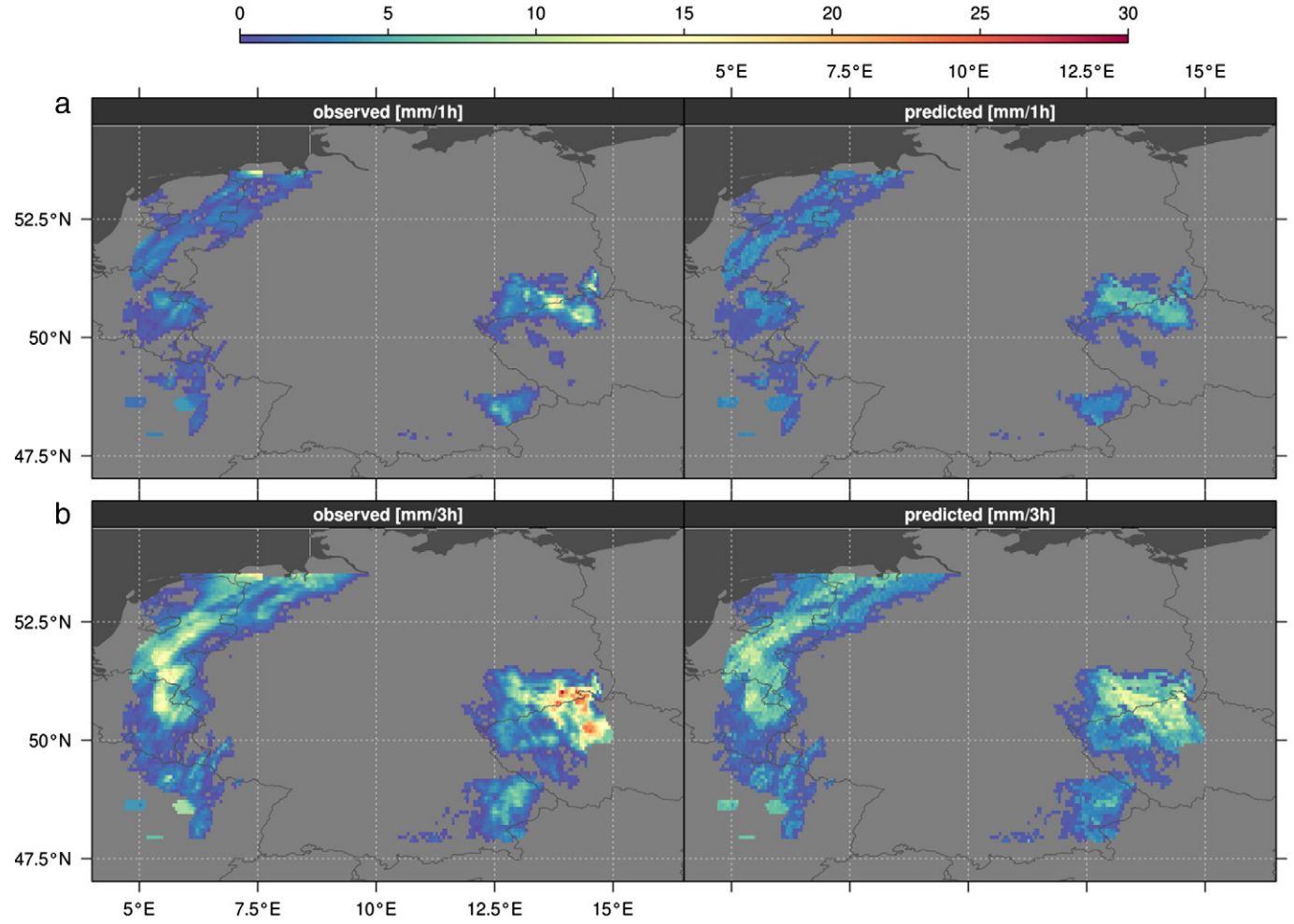

Fig. 5. Rainfall rates observed by radar (left) and predicted by RF model (right) during nighttime. (a) rainfall rates for scene from 15 August 2010 23:00 UTC, (b) aggregated rainfall rates for scenes from 15 August 2010 23:00 UTC to 16 August 2010 1:00 UTC.

was to capture rainfall rates from both, advective-stratiform and convective precipitating cloud areas. Furthermore, the satellite-based estimates of rainfall were realised during day, night and twilight conditions resulting in a 24-hour prediction.

The final rainfall rate assignment technique was realised in three steps: (i) Precipitating cloud areas are identified. (ii) The precipitating cloud areas were separated into convective and advective-stratiform dominated precipitation areas. (iii) Rainfall rates were assigned to the convective and advective-stratiform dominated precipitation areas, respectively. Since the purpose of this study was to explicitly evaluate the potential of random forests for an improved rainfall rate assignment the rain area and rain process detected by the radar network was taken as basis for the investigation. Considering the dominant precipitation processes of convective and stratiform precipitation areas within extra-tropical cyclones, satellite-based information on the cloud top height, cloud top temperature, cloud phase and cloud water path were chosen as predictor variables. Precipitation events between April and September 2010 were chosen. Because of differing information content about the cloud properties during daytime, night-time and twilight conditions, the data set of precipitating events were split accordingly and treated separately. For the training and validation, the radar-based RADOLAN RW product from the DWD which provide area-wide gauge-adjusted hourly precipitation information, was used.

The development of the rainfall rate retrieval technique was realised in three steps. First, an extensive tuning study was carried out to optimise each of the RF models. Second, the RF models were trained using the optimum values for *ntree* and *mtry* found in the tuning study. Third, the RF models were applied to the validation data sets, respectively. Then the predicted rainfall rates are validated against co-located rainfall rates observed by the radar. The training, as well as prediction, were completed on an hourly basis. The rainfall rates were aggregated and also evaluated against each other for 3-h and 8-h interval.

In order to assess the model performances four different evaluation strategies were employed. First, the overall performance of the rainfall rate assignment technique during daytime, night-time and twilight conditions was investigated. Then, the performance on a scene-by-scene basis was considered closely before the diurnal performance of the RF models was shown. Finally, the influence of different training data sets on the model performances was investigated.

Regarding the overall performance, as indicated by Rsq, hourly RR_{Obs} and RR_{Pred} show already a good correlation with Rsq = 0.5 (day and night) and Rsq = 0.48 (twilight). Higher temporal aggregation leads to better agreement. Rsq rises to 0.78 (day), 0.77 (night) and 0.75 (twilight) for 8-h interval. However, a consistent under-estimation (over-estimation) of the maximum (minimum) is apparent for all sets of models, regardless of aggregation or time of day. This shows that very high and low observed rainfall rates cannot be captured by RF because of the averaging of the individual predictions over all trees. In addition, it is not possible to predict beyond the range of response values in the training data.

Comparing day, night and twilight performance show that daytime precipitation is generally predicted best by the model. Twilight and night-time predictions are generally less accurate but only by a small margin. The most obvious difference between the times of day lies in

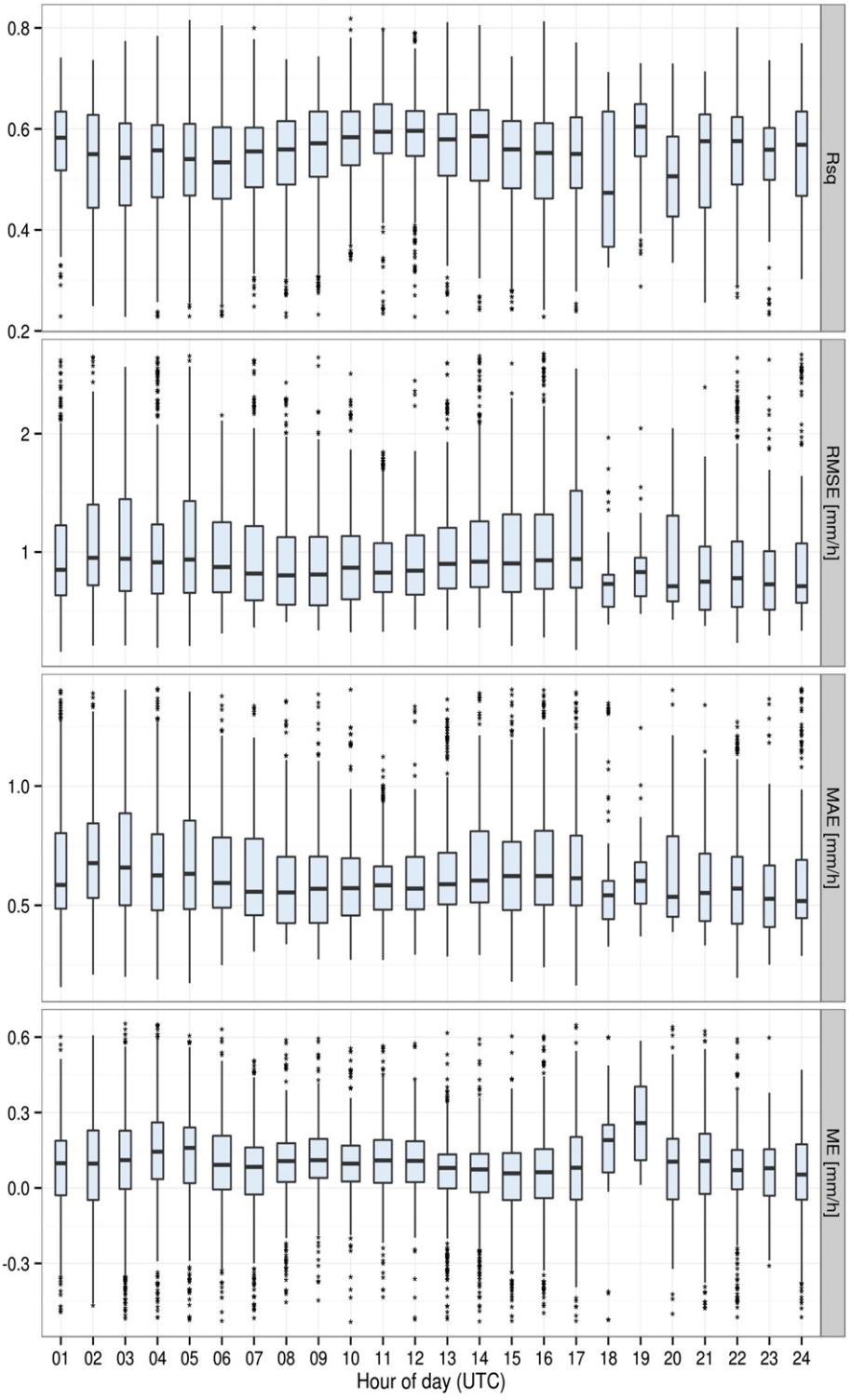

Fig. 6. Diurnal performance of the rainfall rate assignment technique. Box and whisker plots showing distribution of standard verification scores (rows) of RR_{obs} vs. RR_{pred} for each hour of the day. Boxes indicate 25th, 50th and 75th percentiles. Whiskers extend to the most extreme data point within 1.5 times the interquartile range (75th–25th percentiles). Outliers shown as stars. Box widths are relative to number of observations. Extreme outliers (beyond mean +/− 2 times standard deviation) have been removed. Abbreviations are as follows: Rsq, coefficient of determination; RMSE, root mean square error; MAE, mean absolute error; ME, mean error.

the spread of Rsq values. In comparison to daytime scenes, twilight and night-time scenes clearly show greater variability. The smaller number of predictor variables during twilight and night-time conditions as well as more difficult conditions to get cloud parameters during twilight might be a reason. Nevertheless, concerning the considerable problems of existing optical retrievals particularly during twilight, these results reveal a clear and unprecedented improvement and offer the possibility for 24 hour rainfall rate estimation.

The investigation on the influence of different training data sets on the model performances shows that different training data sets do influence the overall prediction accuracy of the according validation data set. Since the observations of the training data sets are used to

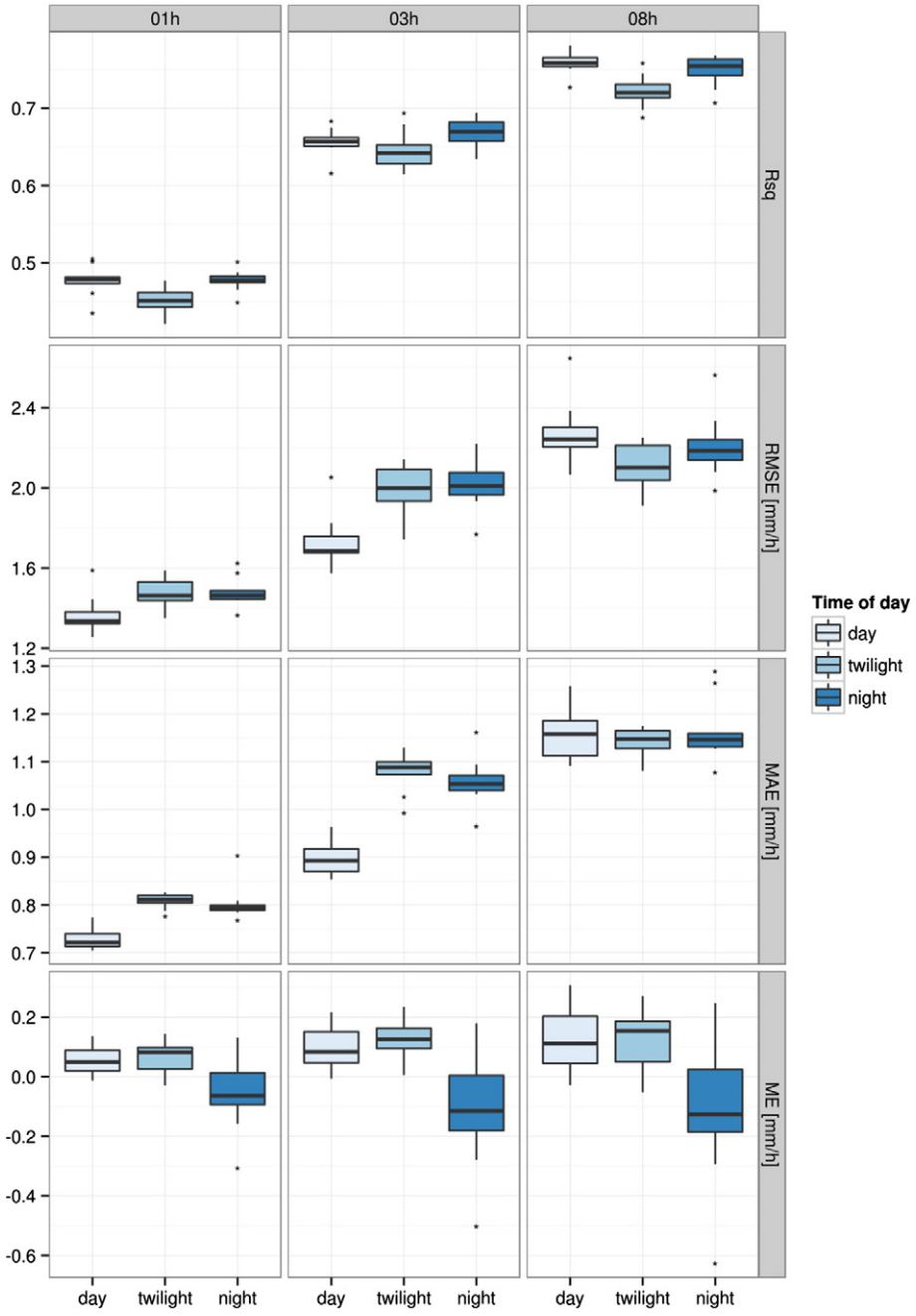

Fig. 7. Influence of different input data sets to the performance of the rainfall rate assignment technique. Box and whisker plots showing distribution of standard verification scores (rows) of RR_{obs} vs. RR_{pred} for different aggregation times (columns) according to time of day (colours). Boxes indicate 25th, 50th and 75th percentiles. Whiskers extend to the most extreme data point within 1.5 times the interquartile range (75th-25th percentiles). Outliers shown as stars. The complete analysis was rerun 10 times, each with different randomly split training and validation sets. Abbreviations are as follows: Rsq, coefficient of determination; RMSE, root mean square error; MAE, mean absolute error; ME, mean error.

build the RF model, these results are not surprising. However, the ranges of the standard verification scores are in a small margin. It seems that the biggest influence is on twilight data. The daytime data sets show the highest median and smallest range of Rsq values for all aggregation levels.

The results show that with the newly developed technique it is possible to assign rainfall rates with good accuracy even on an hourly basis. Furthermore, the rainfall rates can be assigned during day, night and twilight conditions which enables the estimation of rainfall rates for 24 h of a day. This shows great potential for upcoming optical rainfall retrievals. In this context, the spectral resolution provided by MSG SEVIRI offers the possibility for area-wide rainfall rate retrieval in near-real time and in quasi-continuous manner. Furthermore, the potential of rainfall rate assignment based on random forests is confirmed. However, further investigations are necessary to develop a final operational retrieval technique. In the next step, the rainfall rate assignment technique will be combined with a rain area detection and process separation technique. A combined evaluation scheme of precipitation detection, process separation and rainfall rate assignment is intended and absolutely necessary. The complete rainfall retrieval technique will be extensively validated against radar-based RADOLAN RW data. The latter allows a quantification of the total error of the final operational rainfall retrieval technique.

Acknowledgements

The authors are grateful to the German Weather Service (DWD) for providing the RADOLAN and the EUMETSAT Earth Observation Portal (https://eoportal.eumetsat.int/) for providing the Meteosat data set. We would like to thank Jan Cermak from the University of Bochum for providing his cloud mask algorithm, Bernhad Seeger from the University of Marburg for the fruitful co-operation in the context of the Meteosat processing scheme and Tobias Ebert and Johannes Drönner for an excellent implementation. Finally, we thank the members of the Ecosystem Informatics PhD program funded by the University of Marburg for the valuable discussions. The work of M. Kühnlein has partly been supported by the Priority Program 1374 "Biodiversity Exploratories" and the work of T. Appelhans by the Research Unit 1246 "Kilimanjaro ecosystems under global change", both funded by the German Research Foundation (DFG), see www.environmentalinformatics-marburg.de.

References

Ackermann, S. A., Strabala, K. I., Menzel, W. P., Frey, R. A., Moeller, C. C., & Gumley, L. E. (1998). Discriminating clear sky from clouds with MODIS. *Journal of Geophysical Research, 103*(D24), 32141–32157. http://dx.doi.org/10.1029/1998JD200032.

Adler, R. F., Kidd, C., Petty, G., Morissey, M., & Goodman, H. M. (2001). Intercomparison of global precipitation products: The third Precipitation Intercomparison Project (PIP-3). *Bulletin of the American Meteorological Society*, *82*(7), 1377–1396 (Retrieved from http://adsabs.harvard.edu/abs/2001BAMS…82.1377)

Adler, R. F., & Mack, R. A. (1984). Thunderstorm cloud height–rainfall rate relations for use with satellite rainfall estimation techniques. *Journal of Climate and Applied Meteorology*, *23*, 280–296.

Adler, R. F., & Negri, A. J. (1988). A satellite infrared technique to estimate tropical convective and stratiform rainfall. *Journal of Applied Meteorology*, *27*, 30–51.

Aminou, D.M.A. (2002). MSG's SEVIRI instrument. *ESA Bulletin*, *111*, 15–17.

Amorati, R., Alberoni, P. P., Levizzani, V., & Nanni, S. (2000). IR-based satellite and radar rainfall estimates of convective storms over northern Italy. *Meteorological Applications*, *7*(1), 1–18. http://dx.doi.org/10.1017/S1350482700001328.

Arkin, P. A., & Meisner, B. N. (1987). The relationship between large-scale convective rainfall and cold cloud over the western hemisphere during 1982 –84. *Monthly Weather Review*, *115*, 51–74.

Ba, M. B., & Gruber, A. (2001). GOES multispectral rainfall algorithm (GMSRA). *Journal of Applied Meteorology*, *40*, 1500–1514. http://dx.doi.org/10.1175/15200450(2001)040<1500:GMRAG>2.0.CO;2.

Bartels, H., Weigl, E., Reich, T., Lang, P., Wagner, A., & Kohler, O. (2004). *Projekt RADOLAN — Routineverfahren zur Online-Aneichung der Radarniederschlagsdaten mit Hilfe von automatischen Bodenniederschlagsstationen (Ombrometer), 1–111.*

Baum, B.A., & Platnick, S. (2006). Introduction to MODIS cloud products. *Earth Science Satellite Remote Sensing* (pp. 87–108).

Baum, B.A., Soulen, P. F., Strabala, K. I., King, M.D., Ackerman, S. A., & Menzel, W. P. (2000). Remote sensing of cloud properties using MODIS airborne simulator imagery during SUCCESS 2. Cloud thermodynamic phase. *Journal of Geophysical Research*, *105*(D9), 11781–11792. http://dx.doi.org/10.1029/1999JD901090.

Boulesteix, A. -L., Janitza, S., Kruppa, J., & König, I. R. (2012). *Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics.* (Munich).

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*(2), 123–140.

Breiman, L. (2001). Random forest. *Machine Learning*, *45*(1), 5–32. http://dx.doi.org/10.1023/A:1010933404324.

Breiman, L., & Cutler, A. (2008). Random forests — classification manual. Retrieved from http://www.stat.berkeley.edu/breiman/RandomForests/cc_home.htm

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees.* Monterey, CA: Wadsworth and Brooks.

Briem, G. (2002). Multiple classifiers applied to multisource remote sensing data. *IEEE Transactions on Geoscience and Remote Sensing*, *40*(10), 2291–2299 (Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1105916)

Bylander, T. (2002). Estimating generalization error on two-class datasets using out-of-bag estimates. *Machine Learning*, *48*, 287–297 (Retrieved from http://www.springerlink.com/index/JUPCUXF957MC6TQD.pdf)

Capacci, D., & Conway, B. J. (2005). Delineation of precipitation areas from MODIS visible and infrared imagery with artificial neural networks. *Meteorological Applications*, *12*(04), 291–305. http://dx.doi.org/10.1017/S1350482705001787.

Cermak, J. (2006). *SOFOS—a new satellite-based operational fog observation scheme.* Philipps-Universiät Marburg (Retrieved from http:/diss/z2006/0149/pdf/djc.pdf).

Cermak, J., & Bendix, J. (2008). A novel approach to fog/low stratus detection using Meteosat 8 data. *Atmospheric Research*, *87*(3–4), 279–292. http://dx.doi.org/10.1016/j.atmosres.2007.11.009.

Chan, J. C. -W., & Paelinckx, D. (2008). Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sensing of Environment*, *112*(6), 2999–3011. http://dx.doi.org/10.1016/j.rse.2008.02.011.

Cheng, M., & Brown, R. (1995). Delineation of precipitation areas by correlation of METEOSAT visible and infrared data with radar data. *Monthly Weather Review*, *123*(9), 2743–2757 (Retrieved from http://cat.inist.fr/?aModele=afficheN&cpsidt=3654)

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., et al. (2007). Random forests for classification in ecology. *Ecology*, *88*(11), 2783–2792 (Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/18051647)

Cutler, A., & Stevens, J. R. (2006). Random forests for microarrays. In A. Kimmel, & O. Brian (Eds.), *Methods in enzymology* (pp. 422–432) (411th ed.). San Diego: Academic Press.

Dietterich, T. (2002). Ensemble learning. In M.A. Arbib (Ed.), *The handbook of brain theory and neural networks* (2nd ed.). Cambridge: The MIT Press (Retrieved from http://www.ai.kun.nl/aicourses/bki212a/materiaal/tr.pdf)

Ebert, E. E., Janowiak, J. E., & Kidd, C. (2007). Comparison of near-real-time precipitation estimates from satellite observations and numerical models. *Bulletin of the American Meteorological Society*, *88*(1), 47–64. http://dx.doi.org/10.1175/BAMS-88-1-47.

EUMETSAT (2013). Meteosat Third Generation (MTG) will see the launch of four new satellites from 2018. Retrieved October 4, 2013 from http://www.eumetsat.int/website/home/Satellites/FutureSatellites/MeteosatThirdGeneration/index.html

Foody, G. M. (1995). Land cover classification by an artificial neural network with ancillary information. *International Journal of Geographical Information Systems*, *9*(5), 527–542.

Friedl, M.A., & Brodley, C. E. (1997). Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment*, *61*(3), 399–409.

Friedl, M.A., Brodley, C. E., & Strahler, A. H. (1999). Maximizing land cover classification accuracies produced by decision trees at continental to global scales. *IEEE Transactions on Geoscience and Remote Sensing*, *37*(2), 969–977. http://dx.doi.org/10.1109/36.752215.

Früh, B., Bendix, J., Nauss, T., Paulat, M., Pfeiffer, A., & Schipper, J. W. (2007). Verification of precipitation from regional climate simulations and remote-sensing observations with respect to ground-based observations in the upper Danube catchment. *Meteorologische Zeitschrift*, *16*(3), 275–293. http://dx.doi.org/10.1127/0941-2948/2007/0210.

Germogenova, T. A. (1963). Some formulas to solve the transfer equation in the plane layer problem. *Spectroscopy of Scattering Media. Academy of Sciences of BSSR.* (pp. 36–41).

Ghimire, B., Rogan, J., & Miller, J. (2010). Contextual land-cover classification: Incorporating spatial dependence in land-cover classification models using random forests and the Getis statistic. *Remote Sensing Letters*, *1*(1), 45–54. http://dx.doi.org/10.1080/01431160903252327.

Goldstein, B.A., Polley, E. C., & Briggs, F. B.S. (2011). Random forests for genetic association studies. *Statistical Applications in Genetics and Molecular Biology*, *10*(1), 32. http://dx.doi.org/10.2202/1544-6115.1691.

Grimes, D., Coppola, E., Verdecchia, M., & Visconti, G. (2003). A neural network approach to real-time rainfall estimation for Africa using satellite data. *Journal of Hydrometeorology*, *4*, 1119–1133.

Guo, L., Chehata, N., Mallet, C., & Boukir, S. (2011). Relevance of airborne lidar and multispectral image data for urban scene classification using Random Forests. *ISPRS Journal of Photogrammetry and Remote Sensing*, *66*(1), 56–66. http://dx.doi.org/10.1016/j.isprsjprs.2010.08.007.

Hansen, M., Dubayah, R., & Defries, R. (1996). Classification trees: An alternative to traditional land cover classifiers. *International Journal of Remote Sensing*, *17*(5), 1075–1081.

Heinemann, G., Reudenbach, C., Heuel, E., Bendix, J., & Winiger, M. (2001). Investigation of summertime convective rainfall in Western Europe based on a synergy of remote sensing data and numerical models. *Meteorology and Atmospheric Physics*, *76*(1–4), 23–41. http://dx.doi.org/10.1007/s007030170037.

Hong, Y., Hsu, K. -L., Sorooshian, S., & Hiaogang, G. (2004). Precipitation estimation from remotely sensed imagery using an artificial neural network cloud classification system. *Journal of Applied Meteorology*, *43*, 1834–1852 (Retrieved from http://journals.ametsoc.org/doi/pdf/10.1175/JAM2173.1)

Houze, R. A., Jr. (1993). *Cloud dynamics.* San Diego: Academic Press.

Hsu, K. K. -L., Gao, X., & Sorooshian, S. (2002). Rainfall estimation using cloud texture classification mapping. *Proceedings of the 1st IPWG Workshop* (pp. 1–6).

Inoue, T. (1985). On the temperature and effective emissivity determination of semi-transparent cirrus clouds by bi-spectral measurements in the 10-μm window region. *Journal of the Meteorological Society of Japan*, *63*, 88–99.

Kidd, C., & Huffman, G. (2011). Review Global precipitation measurement, *353*, 334–353. http://dx.doi.org/10.1002/met.284.

Kidd, C., & Levizzani, V. (2011). Status of satellite precipitation retrievals. *Hydrology and Earth System Sciences*, *15*(4), 1109–1116. http://dx.doi.org/10.5194/hess-15-1109-2011.

King, M.D. (1987). Determination of the scaled optical thickness of clouds from reflected solar radiation measurements. *Journal of the Atmospheric Sciences*, *44*(13), 1734–1751.

Kokhanovsky, A. A., & Nauss, T. (2005). Satellite based retrieval of ice cloud properties using a semi-analytical algortihm. *Journal of Geophysical Research*, *110*(D19), 1–51 (D).

Krogh, A., & Vedelsby, J. (1995). Neural network ensembles, cross validation, and active learning. *Advances in Neural information processing systems*, *7*. (pp. 231–238).

Kühnlein, M., Appelhans, T., Thies, B., Kokhanovsky, A. A., & Nauss, T. (2013). An evaluation of a semi-analytical cloud property retrieval using MSG SEVIRI, MODIS and CloudSat. *Atmospheric Research*, *122*, 111–135. http://dx.doi.org/10.1016/j.atmosres.2012.10.029.

Kühnlein, M., Thies, B., Nauss, T., & Bendix, J. (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. http://dx.doi.org/10.1175/2010JAMC2284.1.

Lensky, I. M., & Rosenfeld, D. (2003). A night-time delineation algorithm for infrared satellite data based on microphysical considerations. *Journal of Applied Meteorology*, *42*(9), 1218–1226.

Levizzani, V. (2003). Satellite rainfall estimates: New perspectives for meteorology and climate from the EURAINSAT project. *Annals of Geophysics*, *46*, 363–372.

Levizzani, V., Porcu, F., & Prodi, F. (1990). Operational rainfall estimation using Meteosat infrared imagery. An application in Italy's Arno river basin. Its potential and drawbacks. *ESA Journal*, *14*, 313–323.

Levizzani, V., Schmetz, J., Lutz, H. J., Kerkmann, J., Alberoni, P. P., & Cervino, M. (2001). Precipitation estimations from geostationary orbit and prospects for Meteosat Second Generation. *Meteorological Applications*, *8*(1), 23–42. http://dx.doi.org/10.1017/S1350482701001037.

Liaw, A., & Wiener, M. (2002). Classification and regression by Random Forest. *R News*, *2*(3), 18–22.

Malley, J.D., Malley, K. G., & Pajevic, S. (2011). *Statistical learning for biomedical data*. Cambridge University Press.

Mas, J. F., & Flores, J. J. (2008). The application of artificial neural networks to the analysis of remotely sensed data. *International Journal of Remote Sensing*, *29*(3), 617–663. http://dx.doi.org/10.1080/01431160701352154.

Minnis, P., Sun-Mack, S., Young, D. F., Heck, P. W., Garber, D. P., & Chen, Y. (2011). CERES edition-2 cloud property retrievals using TRMM VIRS and terra and aqua MODIS data—Part I: Algorithms. *IEEE Transactions on Geoscience and Remote Sensing*, *49*(11), 4374–4400. http://dx.doi.org/10.1109/TGRS.2011.2144601.

Mota, J. F., Jiménez, M. L., Amate, J. J., & Peñas, J. (2002). *Phytogeographical relationships among high mountain areas in the Baetic Ranges (South Spain), 497–504.*

Mountrakis, G., Im, J., & Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *66*(3), 247–259. http://dx.doi.org/10.1016/j.isprsjprs.2010.11.001.

Nauss, T., & Kokhanovsky, A. A. (2006). Discriminating raining from non-raining clouds at mid-latitudes using multispectral satellite data. *Atmospheric Chemistry and Physics*, *6*(1), 5031–5036. http://dx.doi.org/10.5194/acpd-6-1385-2006.

Nauss, T., & Kokhanovsky, A. A. (2007). Assignment of rainfall confidence values using multispectral satellite data at mid-latitudes: First results. *Advances in Geosciences*, *10*, 99–102.

Nauss, T., & Kokhanovsky, A. A. (2011). Retrieval of warm cloud optical properties using simple approximations. *Remote Sensing of Environment*, *115*(6), 1317–1325.

Negri, A. J., & Adler, R. F. (1993). An intercomparison of three satellite infrared rainfall techniques over Japan and surrounding waters. *Journal of Applied Meteorology*, *32*, 357–373.

O'Sullivan, F., Wash, C. H., Stewart, M., & Motell, C. E. (1990). Rain estimation from infrared and visible GOES satellite data. *Journal of Applied Meteorology*, *29*, 209–223.

Ou, S.C., Liou, K. N., Gooch, W. M., & Takano, Y. (1993). Remote sensing of cirrus cloud parameters using advanced very-high-resolution radiometer 3.7- and 10.9-μm channels. *Applied Optics*, *32*(12), 2171–2180 (Retrieved from http://www.sciencedirect.com/science/article/B6WPY-3V4P93J-25Y/2/3a9072b52afd00af05d865e70fe42bf3)

Ou, S.C., Liou, K. N., Takano, Y., Higgins, G. J., Larsen, N., & Slonaker, R. (2002). Cloud effective particle size and cloud optical thickness. *Raytheon Systems*, 195.

Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, *26*(1), 217–222. http://dx.doi.org/10.1080/01431160412331269698.

Platnick, S., King, M.D., Ackerman, S. A., Menzel, W. P., Baum, B.A., Riédi, J. C., et al. (2003). The MODIS cloud products: Algorithms and examples from Terra. *IEEE Transactions on Geoscience and Remote Sensing*, *41*(2), 459–473.

Polonsky, I. N., Labonnote, L. C., & Cooper, S. (2008). Level 2 cloud optical depth product process description and interface control document, version 5.0. CloudSat Project, CIRA. Retrieved from http://www.cloudsat.cira.clostate.edu/ICD/2BTAU/2B-TAU_PDICD_5.0.pdf

Pompei, A., Marrocu, M., Boi, P., & Dalu, G. (1995). Validation of retrieval algoithms for the infrared remote sensing of precipitation with the Sardinian gauge network data. *Il Nuovo Cimento*, *18 C*, 483–496.

Porcu, F., & Levizzani, V. (1992). Cloud classification using METEOSAT VIS-IR imagery. *International Journal of Remote Sensing*, *13*(5), 893–909.

Prasad, A.M., Iverson, L. R., & Liaw, A. (2006). Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *Ecosystems*, *9*(2), 181–199. http://dx.doi.org/10.1007/s10021-005-0054-1.

Prigent, C. (2010). Precipitation retrieval from space: An overview. *Comptes Rendus Geoscience*, *342*(4–5), 380–389. http://dx.doi.org/10.1016/j.crte.2010.01.004.

R Development CoreTeam (2008). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing (Retrieved from http://www.r-project.org)

Rao, N. X., Ou, S.C., & Liou, K. N. (1995). Removal of the solar component in AVHRR 3.7-μm radiances for the retrieval of cirrus cloud parameters. *Journal of Applied Meteorology*, *34*(2), 482–499. http://dx.doi.org/10.1175/1520-0450(1995)034<0482:ROTSCI>2.0.CO;2.

Reudenbach, C. (2003). *Konvektive Sommerniederschläge in Mitteleuropa Eine Kombination aus Satellitenfernerkundung und numerischer Modellierung zur automatischen Erfassung mesoskaliger Niederschlagsfelder*. Bonner Geographische Abhandlungen. Philipps University Marburg.

Reudenbach, C., Nauss, T., & Bendix, J. (2007). Retrieving precipitation with GOES, Meteosat and Terra/MSG at the tropics and mid-latitudes. In V. Levizzani, P. Bauer, & F. J. Turk (Eds.), *Measuring precipitation form space. Advances i, Vol. 88*. (pp. 509–519). Netherlands: Springer.

Rivolta, G., Marzano, F. S., Coppola, E., & Verdecchia, M. (2006). Artificial neural-network technique for precipitation nowcasting from satellite imagery. *Advances in Geosciences*, *7*, 97–103.

Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (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. http://dx.doi.org/10.1016/j.isprsjprs.2011.11.002.

Roebeling, R. A., & Holleman, I. (2009). SEVIRI rainfall retrieval and validation using weather radar observations. *Journal of Geophysical Research*, *114*(D21), 1–13. http://dx.doi.org/10.1029/2009JD012102.

Rosenfeld, D., & Gutman, G. (1994). Retrieving microphysical properties near the tops of potential rain clouds by multispectral analysis of AVHRR data. *Atmospheric Research*, *34*, 259–283.

Rosenfeld, D., & Lensky, I. M. (1998). Satellite-based insights into precipitation formation processes in continental and maritime convective clouds. *Bulletin of the American Meteorological Society*, *79*, 2457–2476. http://dx.doi.org/10.1175/1520-0450(2003)042<1227:SIIPFP>2.0.CO;2.

Ruiz-Gazen, A., & Villa, N. (2007). Storms prediction: Logistic regression vs random forest for unbalanced data. *Case Studies in Business, Industry and Government Statistics*, *1*(2), 91–101 (Retrieved from http://arxiv.org/abs/0804.0650)

Schmetz, J., Pili, P., Tjemkes, S., Just, D., Kerkmann, J., Rota, S., et al. (2002). An introduction to Meteosat second generation (MSG). *American Meteorological Society*, *83*(7), 977–992. http://dx.doi.org/10.1175/BAMS-83-7-Schmetz-1.

Schmetz, J., Tjemkes, S. A., Gube, M., & van de Berg, L. (1997). Monitoring deep convection and convective overshooting with Meteosat. *Advances in Space Research*, *19*(3), 433–441.

Steele, B.M. (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.

Stone, R. S., Stephens, G. L., Platt, C. M. R., & Banks, S. (1990). The remote sensing of thin cirrus cloud using satellites, lidar and radiative transfer theory. *Journal of Applied Meteorology*, *29*(5), 353–366.

Strabala, K. I., Ackerman, S. A., & Menzel, W. P. (1994). Cloud properties inferred from 8–12 μm data. *Journal of Applied Meteorology*, *33*, 212–229.

Strobl, C., Malley, J., & Tutz, G. (2009). An introduction to recursive partitioning: Rationale, application and characteristics of classification and regression trees, bagging and random forests. *Psychological Methods*, *14*(4), 323–348. http://dx.doi.org/10.1037/a0016973.

Thies, B., & Bendix, J. (2011). Review satellite based remote sensing of weather and climate: Recent achievements and future perspectives. *Meteorological Applications*, *295*, 262–295. http://dx.doi.org/10.1002/met.288.

Thies, B., Nauss, T., & Bendix, J. (2008a). Discriminating raining from non-raining cloud areas at mid-latitudes using Meteosat Second Generation SEVIRI nighttime data. *Meteorological Applications*, *15*, 219–230.

Thies, B., Nauss, T., & Bendix, J. (2008b). Discriminating raining from non-raining cloud areas at mid-latitudes using Meteosat Second Generation SEVIRI daytime data. *Atmospheric Chemistry and Physics*, *8*, 2341–2349. http://dx.doi.org/10.1029/22008JD010464.

Thies, B., Nauss, T., & Bendix, J. (2008c). Precipitation process and rainfall intensity differentiation using Meteosat Second Generation SEVIRI data. *Journal of Geophysical Research*, *113*, 19. http://dx.doi.org/10.1029/2008JD010464.

Thies, B., Nauss, T., & Bendix, J. (2008d). First results on a process-oriented rain area classification technique using Meteosat Second Generation SEVIRI night-time data. *Advances in Geosciences*, *8*, 1–9.

Tjemkes, S. A., van de Berg, L., & Schmetz, J. (1997). Warm water vapour pixels over high clouds as observed by Meteosat. *Beiträge zur Physik der Atmosphäre*, *70*(1), 15–21.

Wu, R., Weinman, J. A., & Chin, R. T. (1985). Determination of rainfall rates from GOES satellite images by a pattern recognition technique. *Journal of Atmospheric and Oceanic Technology*, *2*, 314–330.