DISSERTATION

submitted to the

Combined Faculty for the Natural Sciences and Mathematics

of

Heidelberg University, Germany

for the degree of
Doctor of Natural Sciences

Put forward by
Diplom-Informatiker Fabian Elias Bachl
born in Berlin

Date of oral examination: .....

Bayesian Hierarchical Models for Remote Assessment of Atmospheric Dust

Advisors: PD Dr. Christoph S. Garbe
Prof. Tilmann Gneiting

Abstract

Dust storms emerging in the Earth's major desert regions significantly influence weather processes, the CO_2-cycle and the climate on a global scale. Their effects on organisms range from providing nutrition to vegetation and microbes to direct impact on human settlements, transportation and health. The detection of dust storms, the prediction of their development, and the estimation of sources are therefore of immediate interest to a wide range of scientific disciplines. Recent spatio-temporal resolution increases of remote sensing instruments have created new opportunities to understand these phenomena. The scale of the data and their inherent stochasticity, however, pose significant challenges. This thesis develops a combination of methods from statistics, image processing, and physics that paves the way for efficient probabilistic dust assessment using satellite imagery. As a first step, we propose a Bayesian hierarchical model (BHM) that maps Spinning Enhanced Visible and Infrared Imager (SEVIRI) measurements to a predictor of the dust density. Case studies demonstrate that, as compared to linear methods, our latent signal mapping (LSM) approach mitigates effects of signal intrinsic noise on further processing steps. Furthermore, an extensive cross-validation study is employed to show that LSM successfully adapts to intra-daily changes of the infrared data and yields outstanding dust detection accuracy. Physically, the dust density and its transport process are tied together by the continuity equation. A traditional approach to determine the flow field for a given density is the variational method of Horn and Schunck (HS), which simplifies the equation to compression free motion. We characterize the equation's solution as a Gaussian Markov random field (GMRF) and introduce compressible dynamics. This link between probabilistic and variational perspectives leads to applied and theoretical advances. It enables us to employ the integrated nested Laplace approximation (INLA) technique for computationally efficient inference and integration over hyper-parameters. The importance of allowing for compressible motion and treating the problem in a statistical manner is emphasized by simulation and case studies showing a significant reduction in errors of the estimated flow field. In addition, we demonstrate how our methodology provides uncertainty quantification, dust storm forecasts and estimation of emission sources. The thesis is concluded by examining the analytical properties of our approach. It is shown that, under mild restrictions on an underlying Sobolev space, existence and uniqueness of the compressible flow can be guaranteed on a continuous domain and a well-posed discretization exists. Lastly, our variational calculations point to an interpretation of the density as a solution to flow-parameterized stochastic partial differential equations (SPDEs) naturally extending Matérn fields to non-isotropy, which provides a further step towards a joint model of dust density and flow field.

Zusammenfassung

Weltweit entstehen in Wüstenregionen Sandstürme, die einen signifikanten Einfluss auf Wetterprozesse, den CO_2-Zyklus sowie das Klima in globalem Maßstab ausüben. Sie stellen der Vegetation und Organismen lebensnotwendige Nährstoffe zur Verfügung und wirken sich gleichzeitig auf menschliche Siedlingsräume, das Transportwesen sowie die Gesundheit aus. Die Detektion von Sandstürmen, die Vorhersage ihrer Entwicklung und die Ermittlung von Quellregionen sind daher von großem Interesse für ein breites Spektrum wissenschaftlicher Disziplinen. Die aktuelle Progression der raumzeitlichen Auflösung von Instrumenten der Satellitenfernerkundung bietet in dieser Hinsicht neuartige Möglichkeiten, diese Phänomene zu untersuchen. Der Umfang der gewonnenen Daten und ihre inhärente Stochastizität bedeuten jedoch gleichzeitig eine signifikante Herausforderung. In der vorliegenden Dissertation wird ein System zur effizienten probabilistischen Erfassung von Sandstürmen entwickelt, welches Methoden der Statistik, Bildverarbeitung und Physik miteinander vereint. Zunächst wird ein Bayesianisches hierarchisches Modell (BHM) vorgestellt, welches Spinning Enhanced Visible and Infrared Imager (SEVIRI) Messungen auf einen Prädiktor der Staubdichte abbildet. Anhand von Fallstudien zeigen wir auf, dass unser latent signal mapping (LSM) Ansatz im Vergleich zu linearen Methoden durch Rauschunterdrückung darauf aufbauende Verarbeitungsschritte erleichtert. Mittels einer umfangreichen Kreuzvalidierungsstudie wird zudem die herausragende Detektionsgenauigkeit der LSM-Methode in Anbetracht von täglichen Schwankungen des Infrarotsignals belegt. Aus physikalischer Sicht wird der Zusammenhang zwischen der Staubdichte und ihrem Transportprozess durch die Kontinuitätsgleichung beschrieben. Ein traditioneller Ansatz der Variationsrechnung zur Bestimmung des inkompressiblen Flusses aus gegebener Dichte ist die Methode von Horn und Schunck (HS). Wir charakterisieren die Lösung der Gleichung als ein Gaussian Markov random field (GMRF) und integrieren kompressible Flüsse in das Modell. Diese Verbindung zwischen Probabilistischer Modellierung und Variationsrechnung führt aus Sicht der Anwendung sowie Theorie zu weitreichenden Fortschritten. Mittels der integrated nested Laplace approximation (INLA) Methode erlaubt sie effiziente computergestützte Integration über Modellparameter. Die bedeutende Rolle dieses Vorgehens sowie der Kompressibilität demonstrieren wir mittels einer Simulationsstudie, welche eine signifikante Reduktion von Fehlern im ermittelten Flussfeld nachweist. Des Weiteren zeigen wir, wie unsere Methode Unsicherheit quantifiziert, der Vorhersage von Sandstürmen dient und zur Ermittlung von Emissionsquellen eingesetzt werden kann. Abschließend werden die analytischen Eigenschaften unseres Ansatzes vertiefend untersucht. Unter der Annahme eines schwach eingeschränkten Sobolev-Raumes zeigen wir Existenz und Eindeutigkeit des kompressiblen Flusses über einem kontinuierlichen Definitionsbereich sowie einer wohldefinierten Diskretisierung. Durch Variationsrechnung wird zudem eine Verbindung zu einer Klasse von stochastischen Differentialgleichungen (SPDEs) aufgezeigt, welche Matérn-Felder anisotrop fortsetzen. Ihre Parametrisierung durch Flussfelder und Lösung durch Materialdichten bietet einen weiteren Schritt in Richtung probabilistischer Modellierung der gemeinsamen Verteilung von Staubdichte und Bewegung.

Acknowledgements

During my time as a Ph. D. student I was supported by many great people. Christoph S. Garbe, my first supervisor, provided me with a great amount of freedom in pursuing my scientific endeavors. He always had a sympathetic ear for professional topics and, most importantly, for private matter as well. Not being a statistician it was a delighting experience to work with Tilmann Gneiting, my second supervisor, Thordis L. Thorarinsdottir and Alex Lenkoski. Their support and dedication provided major contributions to my interdisciplinary project as well as my personal development as a scientist. I would also like to thank Christoph Schnörr, the head of my research training group. Not only did he make our mathematical research one of the most enjoyable parts of my years as a Ph. D. student but would also always provide fruitful advice, structuring discussions and an magnificent working environment in general.

Last but not least my gratitude goes to my friends, family and colleagues. If I would mention your names and what you mean to me this thesis would need another 50 pages. Your support and confidence is still and always has been incredible and I am the luckiest man on earth to share my life with you.

Rest in peace, Opa Bachl.

Funding by the Deutsche Forschungsgemeinschaft via the research training group 1653 *Spatio / Temporal Graphical Models and Applications in Image Analysis* is also gratefully acknowledged.

Contents

List of Publications

1. Fabian E. Bachl and Christoph S. Garbe. Classifying and Tracking Dust Plumes from Passive Remote Sensing. In *Proceedings of the ESA, SOLAS & EGU Joint Conference 'Earth Observation for Ocean-Atmosphere Interaction Science'*, pages S1-3, 2012.

2. Fabian E. Bachl, Paul Fieguth and Christoph S. Garbe. A Bayesian Approach to Spaceborn Hyperspectral Optical Flow Estimation on Dust Aerosols In *Proceedings of the International Geoscience and Remote Sensing Symposium 2012*, pages 256–259, 2012.

3. Fabian E. Bachl, Paul Fieguth and Christoph S. Garbe. Bayesian Inference on Integrated Continuity Fluid Flows and their Application to Dust Aerosols In *Proceedings of the International Geoscience and Remote Sensing Symposium 2013*, pages 2246–2249, 2013.

4. Fabian E. Bachl, Alex Lenkoski, Thordis. L. and Christoph S. Garbe. Bayesian Motion Estimation for Dust Aerosols In *ArXiv e-prints arXiv:1308.0469, submitted to Annals of Applied Statistics*, under review, 2013.

List of Acronyms

AAI absorbing aerosol index

AERONET Aerosol Robotic Network

AOD aerosol optical depth

BHM Bayesian hierarchical model

BCE brightness constancy equation

BT brightness temperature

CAR conditional auto-regression

DAG directed acyclic graph

DGM directed graphical model

EHM empirical hierarchical model

ESA European Space Agency

EUMETSAT European Organisation for the Exploitation of Meteorological Satellites

GM graphical model

GMRF Gaussian Markov random field

GLM generalized linear model

HM hierarchical model

HS Horn & Schunck method for optical flow

ICE integrated continuity equation

INLA integrated nested Laplace approximation

IS importance sampling

LDA linear discriminant analysis

LPF latent projection function

LSM latent signal mapping

MCMC Markov chain Monte Carlo

MODIS Moderate Resolution Imaging Spectroradiometer

MRF Markov random field

MSG Meteosat Second Generation

NASA National Aeronautics and Space Administration

PDE partial differential equation

PGM probabilistic graphical model

SEVIRI Spinning Enhanced Visible and Infrared Imager

SFI SEVIRI falsecolor imagery

SPD symmetric positive definite

SPDE stochastic partial differential equation

List of Figures

1 Introduction

1.1 Motivation

Dust storms are global meteorological phenomena originating from arid and semi-arid regions all over the earth. Strong winds blow loose sand and dirt from the ground and carry them into the atmosphere where they are transported for thousands of kilometers and ultimately deposited worldwide. These mixtures of atmospheric gases and particulate matter are called aerosols and interact with other atmospheric components, oceanic and terrestrial chemistry and physics as well as living organisms ranging from microbes and vegetation to human beings.

One immediate effect can be intuited from Figure 1.1 showing a dust plume over northern Africa. Layers of atmospheric dust block the solar radiation from reaching the surface of the earth and cause the latter to cool down[1]. The atmosphere, however, undergoes the opposite effect as more radiative energy is absorbed than under pristine sky conditions. This change of (thermal) energy distribution in turn leads to changes in vertical and advective atmospheric transport processes, which stabilize the atmosphere (Heinold et al., 2008). In addition to this direct impact on the earth's radiation budget and atmosphere circulation there are also significant indirect effects. Dust particles interact with water clouds on a microphysical basis. Thereby cloud formation processes are altered and optical properties are changed, which directly influence weather conditions like precipitation rates (Rosenfeld et al., 2001). Local dust event detection and spatial estimation of its strength days after its genesis is thus of two-fold importance in this context. Integration of such estimates in respective models can yield improved accuracy in weather forecasts. Most importantly, their analysis conveys additional insight into the aforementioned climate processes to environmental scientist.

In addition, human modes of living are affected on a short and long term basis. As dust filters radiation, the respective reduction in range of vision causes road and airport closures. In particular, human settlements that are in the vicinity of desert regions have great interest in immediate detection as well as short and medium range spatial forecasts of dust events. Even more concerning is the effect on human health. Some dust particles are small enough to surpass the natural human defense system and can cause illnesses like asthma by causing damage to the respiratory system

[1] If not stated otherwise, all Figures in this thesis the author's work.

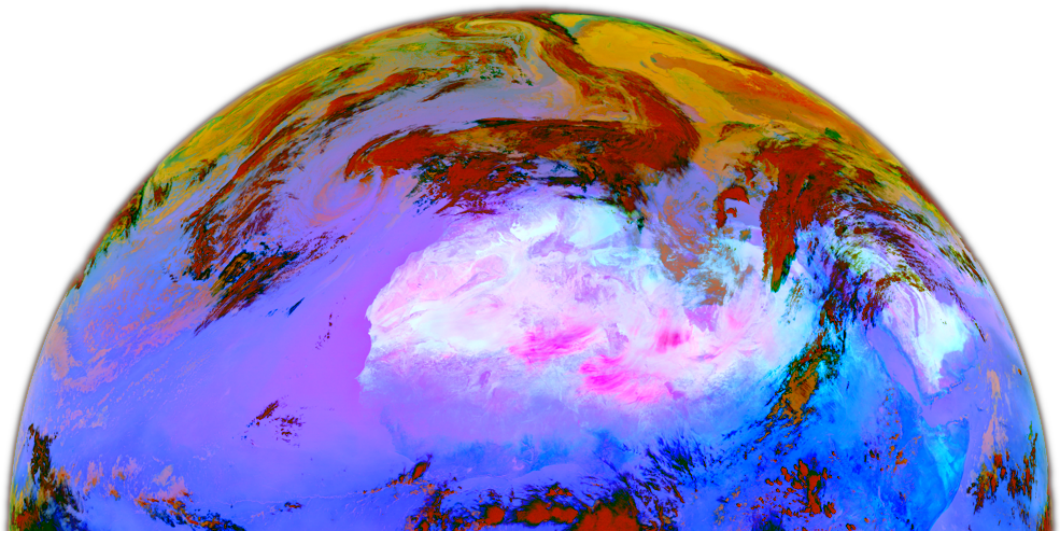
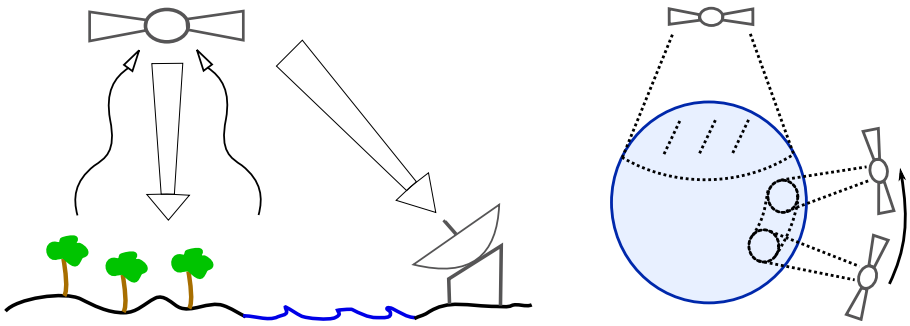


Figure 1.1: Northern hemisphere of the earth showing a dust storm on January 18, 2010 at 12h GMT over northern Africa (pink area). Blue to cyan areas show the surface of the earth under pristine sky conditions. Black and nearby orange to red regions reflect the presence of ordinary water clouds.

(Griffin and Kellogg, 2004). Moreover, bacteria and fungal spores travel with the dust and are suspected to be connected to the so Valley fever in the San Joaquin Valley in California and severe meningitis epidemics between Senegal and Ethiopia (Sultan et al., 2005).

Marine and terrestrial ecosystems are a third factor that underlines the importance of an exact understanding of aerosol genesis and deposition. Minerals that are transported with the dust act as a nutrient to a variety of life forms. Their mineralogical composition is identical to that of the source region the dust was emitted from and hence determines their nutritional properties. For instance, iron makes up about 4% of northern African soil and is at the same time an essential prerequisite to the open ocean phytoplankton production. Changes in dust composition and hence iron fluxes can therefore lead to species shifts. These again have an effect on the global climate as oceanic CO_2 uptake is changed by the efficiency of organic carbon export to deep water (Jickells et al., 2005).

The detection of dust storms, the prediction of their development, and the localization of sources are therefore of immediate interest for a wide range of environmental applications. During the last decades space-born remote sensing systems have gained an indispensable role in this context. By capturing propagated signals like gravitation and electromagnetic radiation from the earth's surface, objects in the atmosphere as well as the ocean they allow for a spatial and temporal range of environmental information retrieval that is not feasible with earth bound in situ measurements.



(a) (b)

Figure 1.2: Spaceborn remote sensing. Panel (a): Passive sensors capture signals emission of which is self induced by the object that is probed. Active instruments emit a signal and record the object's response to this probing, e.g., the reflection of an emitted electromagnetic beam. Satellites are either non-stationary or geostationary, as shown in Panel (b). Geostationary satellites do not move relative to the point of the earth's surface that is closest (nadir) and thus have a constant field of view in terms of probing data. Non-stationary satellites change their location relative to the earth's surface. Their trajectory leads to a variable field of view, called the swath of the satellite.

The respective measurement techniques and instruments can be divided into active and passive approaches (see Figure 1.2(a)). In active remote sensing a signal is emitted from the satellite and the response is recorded. For instance, laser altimeters determine local elevations like ocean waves by measuring the time a light beam needs to travel to the water surface and return to the satellite. From this data wind speeds and direction can be estimated. Passive techniques do not probe by recording the response to a signal but measure signals self-induced by object of interest.

In this thesis data gathered by a passive remote sensing instrument aboard the Meteosat Second Generation (MSG) 9 platform is analyzed. Meteosat, short for Meteorological satellite, is a series of meteorological platforms operated in a cooperation between the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) and the European Space Agency (ESA). The SEVIRI instrument aboard MSG-9 poses an unique opportunity for the analysis of local and sub-daily as well as large scale long term processes of dust emission and transport for multiple reasons.

Infrared measurements do not depend on the illumination provided by the sun and indicate dust events spanning several days even during night. Since MSG platforms are stationary, it is also guaranteed that the spatial and temporal imagery coverage is permanent and a dust event can not be overlooked due to being outside the swath of

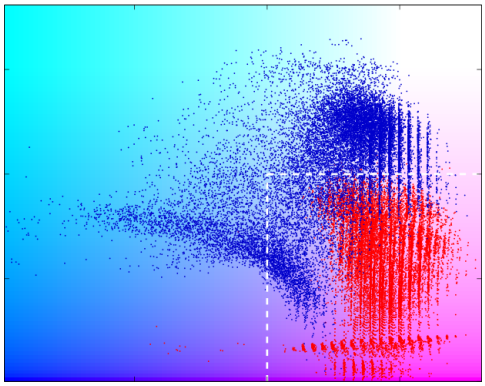
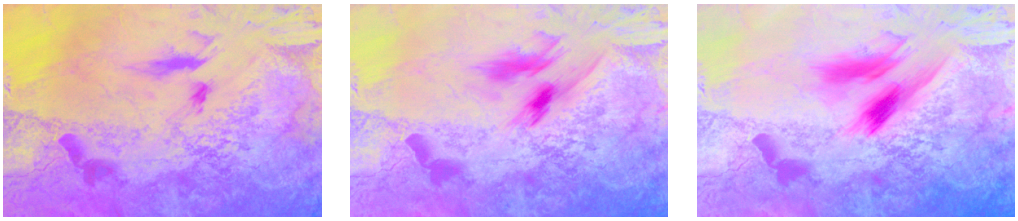


Figure 1.3: Distribution of dusty (red markers) and pristine sky (blue markers) pixels in the color space of the SEVIRI falsecolor imagery. The abscissa and ordinate are the red and green channel intensities. The background color is determined according to these axis and a fully saturated blue channel.

the satellite (see Figure 1.2(b) for an explanation of geostationarity). Most important, however, is the resolution of SEVIRI. At nadir (the earth's surface point that is closest to the satellite) it provides a spatial resolution of 3×3 km decreasing with increasing distance from that point. These measurements are taken every 15 minutes at 12 distinct frequency windows ranging from visible spectra to infrared radiation within 3.9 to 13.4 μm. Figure 1.1 shows this data depicted in falsecolor scheme. The color channels represent differences of frequency specific infrared radiation. As discussed earlier, dust absorbs this radiation depending on the frequency such that the differences represent its presence. Falsecolor imagery is a common mode of visual dust aerosol assessment by scientists and forms the basis of the data analysis carried out throughout this document.

Although SEVIRI's properties in principle lend themselves towards increased insight into aerosol processes the amount of the data collected and the inherent stochasticity pose significant challenges.

A main question that arises is whether a particular pixel in a given SEVIRI image represents a region that is affected by dust presence. At first, when considering Figure 1.1, visual inspection suggests that this might be deemed as an easy task since the pink color representing dust is easily identified by the human eye. Evidence that this is a misleading perception can, however, be seen in the same figure. A large region in north-western Africa that lies mostly inside Morocco appears slightly pink as well, which is caused by the radiative properties of the local ground conditions but not by dust. In Figure 1.3 this issue is quantified by showing a distribution of dusty and pristine sky pixel samples in the color space. Is is easy to see that a simple threshold in one or both of the falsecolor channels is not sufficient for a discrimination of the two kinds.



(a) (b) (c)

Figure 1.4: Intradaily signal variability of SEVIRI. Panels (a) to (c) visualize the development of a dust plume emerging in northern Chad on January 18, 2010 at 7.30 am, 8.30 am and 9.30 am GMT, respectively. A temporal change of the data can also be seen in pristine region, e.g. the loss of yellow tone in the left upper corner and an increase of blue intensity in the right bottom corner.

Figure 1.4 gives an idea about another intricacy dust detection comes with. Among atmospheric changes like water clouds that may visually cover dust plumes the intradaily thermal energy change of the surface due to solar radiation is the most significant contributor to changes in the SEVIRI signal. These and other stochastic influences lead to the insight that a probabilistic model is required to characterize dust presence in the signal. The first thread of this thesis is dedicated to this issue: Bayesian hierarchical models for dust detection.

Once an indication scheme for dust presence available, the second main question arises. In order to mathematically describe the genesis, transport and deposition of dust aerosols characterization of their flow through the atmosphere is required. As this flow is mostly determined by atmospheric winds a readily available approach is to employ wind field estimates derived for weather forecasts and climate analysis. Unfortunately, these fields are usually determined on a spatial scale that is too coarse to describe small scale variabilities of dust events, especially during genesis. A second factor is that the vertical extent of a dust plume given in the imagery is not known. It is hence not clear which atmospheric layer and corresponding wind field estimate to assume to be correct.

An alternative approach, the one that is pursued throughout this thesis, is to infer the flow field of dust aerosols directly from the previously determined dust indicator. The general intuition behind the strategy is to employ the local gradient information of the indicator to obtain a dense estimate of the flow field under the assumption that this field is smooth. In Figure 1.5 such a flow field is depicted together with the corresponding dust indicator. This way rise is given to further predictions with respect to the behavior of the aerosol. As opposed to studies where this was done manually by experts the flow field can now serve as a mean to identify regions that initially emit the dust by a simulated back-transport of a given plume. Furthermore,

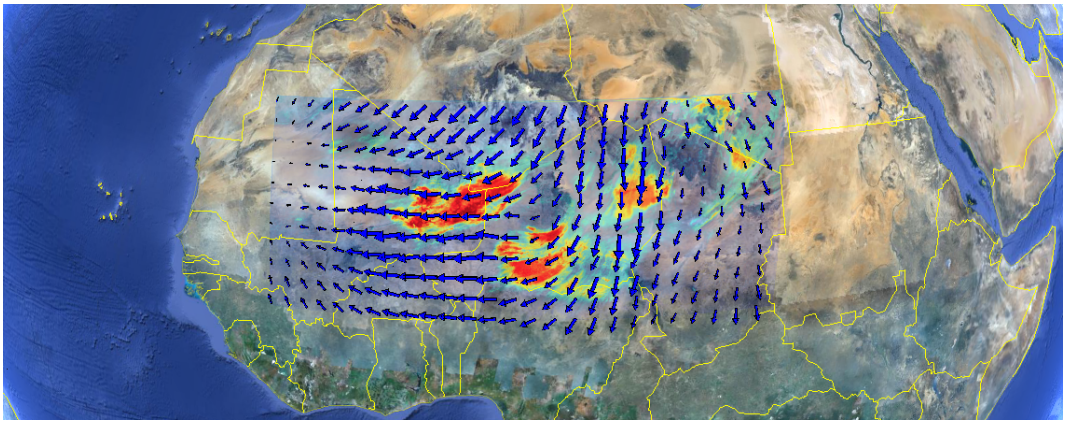


Figure 1.5: Flow field of a dust event over northern Africa on January 17, 2010 at 12h GMT.

the same idea can be employed to simulate the future development of a dust plume and thus give the forecasts that are so highly desirable for the abovementioned reasons.

1.2 Related Work

Contemporary detection schemes for dust aerosols follow two different paradigms. Motivated by physical models of conditions for dust emission, transport via wind fields and radiative filtering properties of aerosols, the work of Klüser and Schepanski (2009), and Brindley et al. (2012) is based on connections between falsecolor imagery and aerosol optical depth (AOD). Here, the presence of dust is quantified by a combination of different thresholds derived from case- and simulation-studies. Two of these proposed thresholds are depicted in Figure 1.3. In contrast, the work of Rivas-Perea et al. (2010) and Eissa et al. (2012) employs methods from machine learning and image processing by using neural nets to learn non-linear dust detection criteria from a data set with labels set by a human expert.

From a statistical viewpoint, both approaches suffer from shortcomings. Directly imposing thresholds partly based on physical assumptions, simulations studies and qualitative inspection of the data by an expert might lead to misleading conclusions if premises are not met or due to human subjectivity. Also, neither Klüser and Schepanski (2009) nor Brindley et al. (2012) include quantification of uncertainty in their analysis. On the other hand, even though interpretable in a probabilistic sense and directly driven by data, the inner mechanics of image processing tools like neural nets are not physically motivated and intransparent to the user. Especially in the context of the environmental sciences this hinders such an approach to further elevate

the understanding of the data. Alongside the purely local detection or quantification of dust of a second objective of concern has, to the best of our knowledge, not yet been addressed.

Further, none of the previously mentioned approaches imposes or performs inference on a coherent spatio-temporal structure like an underlying aerosol transport process. This omits valuable information. Previous attempts to localize and characterize areas being sources of dust storms by Schepanski et al. (2012) have to rely on human visual data inspection and indication or simple temporal averaging. Alternatively, Alonso-Pérez et al. (2012) employ Lagrangian trajectories derived from wind field averages to infer the source of a given in-situ sample.

However, various approaches in different scientific fields capture problems that are related to the task at hand. Statistical approaches are predominantly driven by applications related to either the verification of numerical weather predictions or the issuing of nowcasts, forecasts for very short lead-times, see e.g. Gilleland et al. (2010) and Xu et al. (2005). Here, a transformation between two spatial fields (e.g. a prediction and the corresponding observation) is determined via a deformation field that associates spatial locations of the two fields in a smooth fashion. In prediction problems, the deformation field then serves as a tool to assess the field both in terms of mis-localization and quantification error. In contrast, in nowcasting a current spatial observation and a given deformation field are utilized to predict the spatial field representing future realizations.

Xu et al. (2005) apply a integro-difference equation where information is propagated between the two fields through a kernel function. In image processing differential approaches–which can be interpreted as special cases of the integro-difference equation–have been popular since the advent of the Horn & Schunck method for optical flow (HS) (Horn and Schunck, 1981). The latter derives a displacement field between two images, e.g. in a video sequence, and has two appealing properties. A treatment in variational frameworks is straight forward due to its simple analytic properties. For instance, Schnörr (1991) shows existence and uniqueness properties of the HS optical flow under the assumptions of a mildly restricted underlying Sobolev space. Corpetti et al. (2002) point out that the assumptions underlying the HS approach are a special case of the integrated continuity equation, a physical model of matter transport.

These methods have only entered the statistics community to some extent, see e.g. Marzban and Sandgathe (2010) who employ the connate optical flow approach of Lucas and Kanade (1981). A coherent probabilistic framework is therefore missing although recent statistical research clearly indicates a respective tendency. Lindgren et al. (2011) show a connection between Gaussian fields and GMRFs via SPDEs (see also Simpson et al. (2012)). They employ SPDEs to set forth the concept of Matérn covariances for Gaussian fields that Guttorp and Gneiting (2006) conclude to naturally appear in a number of scientific fields related to environmental research. Most

interestingly, Lindgren et al. (2011) also point out a respective interpretation of the image warping method of Sampson and Guttorp (1992), which itself is closely related to the method of optical flow. In this context, the flow field in fact parameterizes the SPDE and thereby defines the spatio-temporal coherence model imposed on the field that is performed inference on and would, in case of aerosols, model the intensity of the dust itself. Probabilistic inference on the parameters of the SPDE, i.e. the flow field, on the other hand is not treated in these publications and thus remains an open problem.

1.3 Contribution

This thesis and the associated publications contribute to the scientific research with respect to the detection, flow estimation, forecasting and source estimation of dust aerosols from remote sensing data. In particular, a strong emphasis is put on the development and definition of probabilistic graphical models that at the same time capture the stochastic properties of these processes and reflect physically meaningful quantities and interrelationship while maintaining reasonable computational complexity.

The first main topic of this thesis is the detection of dust aerosols from SEVIRI data. Brindley et al. (2012) show that the appearance of the SEVIRI falsecolor data representation described by Lensky and Rosenfeld (2008) is correlated with the AOD at a particular wavelength and propose a thresholding scheme to flag pixels containing dust activity. An initial step towards a respective probabilistic formulation is the logit regression model elaborated on in chapter 3 and introduced by Bachl and Garbe (2012). By introducing spatially varying random effects that depend on an estimate of the earth's infrared radiation under clear sky conditions (further on called background estimate) this model is flexible enough to adapt for non-linear dust presence contributions to the SEVIRI falsecolor imagery (SFI) that are not covered by linear discriminant analysis (LDA) and related methods.

This detection method is then extended to adapt for a problem that occurs when employing its linear predictor for later tasks like the estimation of dust flow. As shown in Bachl et al. (2012) the intrinsic linear projection of the approach directly passes on signal noise of the falsecolor imagery, e.g., atmospheric disturbances. The remedy chosen here is to waive the linear projection by interpreting the signal as one of two domains of the functions contributing to the linear predictor rather than as their multiplicative factor. Due to the imposed regularity assumptions for these random functions this leads to a significantly increased signal to noise ratio with respect to the linear predictor.

Lastly, as pointed out by Bachl et al. (2013a), background estimation is an impeding aspect of these approaches as the respective selection criterion is only applicable at

day time. This is critical as dust plume genesis often takes place in the early morning. It is shown, however, that a monthly average of surface emissivity estimates can serve the same purpose as background estimates. Moreover, with these at hand, dust detection with outstandingly high levels of specificity and sensitivity are feasible.

The second main topic is the estimation of aerosol flow. For this purpose, the HS method is analyzed and it is shown how to formulate this approach as a Bayesian hierarchical model. As initially elaborated in Bachl and Garbe (2012) this gives a probabilistic interpretation of the optical flow as a latent GMRF. While the link is relatively straightforward, to the best of the author's knowledge, this is the first time that the full distributional aspects and the associated uncertainty are taken into account for the HS method. The intrinsic smoothness parameter of the method then finds a clear meaning as the precision hyper-parameter of the conditional autoregression model imposed on the flow. This perspective comes with several long and short term benefits. Firstly, inference can be performed using computationally efficient integrated nested Laplace approximations (INLA) (Rue et al., 2009). A second benefit is the interpretability of the flow field in terms of the physical nature of the phenomenon under consideration. Even in case the HS method is elevated to more robust error penalties this allows for computationally efficient and precise assessment of the posterior flow and its incorporated uncertainties as well as for the hyperparameters.

A second contribution to flow estimation is to leverage the hierarchical Bayesian framework to overcome deficiencies in the HS formulation (Bachl et al., 2013a,b). A typical quirk of statistical warping and optical flow is the underlying preservation assumption of the respective quantity along its trajectory. In dust aerosols (as well as other natural phenomenon) this might lead to false conclusions. Gaseous solutions are compressible and remote sensing often only leads to a non-bijective mapping of a three dimensional quantity to a two dimensional data space. Alongside advection, observations are therefore clearly prone to convective effects resulting from compression of the solution or material exchange inside a projected atmospheric column. As a remedy we extend the HS method to incorporate the water vapor related work of Corpetti et al. (2002) and put it in a Bayesian hierarchical model context. As our work shows in a simulation study, the integrated continuity equation (ICE) considerably reduces errors in the estimated flow field. The main advantage of the ICE comes from the fact that it implicitly considers a multiplicative convective effect that is driven by the divergence of the flow field itself. A motion trajectory starting at a point where the divergence is positive (negative) leads to a low (high) multiplicative effect mimicking the dispersion (accumulation) of the modeled quantity. In case of pure advection, i.e. the absence of divergence, the multiplicative factor is 1 and the usual preservation assumption is retained.

Lastly, it is elucidated how the variational framework popular in image processing can be employed to gain guarantees in terms of the existence and uniqueness of the continuous solution of the ICE energy functional within a mildly restricted Sobolev

space. This implies that the Ritz method for discretization of this problem becomes a finite element formulation for which the solution is equivalent to mean inference on the (now proper) latent GMRF employed for the discrete ICE approach. Moreover, it is shown that if the motion field is held fixed, the Euler-Lagrange equations with respect to the moving matter reveal a connection to the approach SPDE of Lindgren et al. (2011).

1.4 Organization

This thesis is organized as follows. Chapter 2 begins with an introduction to the basic objects and terminology of probability as well as graph theory. With these powerful mathematical tools at hand, probabilistic graphical models (PGMs) are defined and categorized into directed and undirected modes of expressing relationships among random variables. A precise overview is given of how properties of the respective graphs translate into probabilistic independence assumptions and vice versa. As a first example for such models BHMs enter the focus of the chapter. These play a significant role throughout this thesis by providing a consistent framework to impose structure onto the variables interest: observed data, an unknown (latent) process that generates these observations as well as parameters of the process itself. Consecutively, generalized linear models (GLMs) and GMRFs receive the attention of this manuscript. GLMs describe the class of probabilistic dependencies between the observed data and latent processes that are assumed to hold throughout the experiments conducted within this work. Thereby a perspective is taken that describes how the latent process generates the observations from other kinds of data. The properties in terms of prior assumptions about these processes are expressed by the distributional class of GMRFs for multiple reasons. GMRF naturally emerge from particular physical mechanisms, imply a desirable amount of uncertainty where no information from observations is available and provide a computationally exceptional convenient way for discrete approximations to a broad class of continuous models. Chapter 2 is concluded by a comprehensive analysis of INLAs. This technique provides the algorithmic basis for probabilistic inference within the class of aforementioned models and hence the conducted experiments.

Chapter 3 is dedicated to the detection of atmospheric dust using SEVIRI measurements. The data itself and related methods for dust detection like thresholding concepts and LDA are introduced in Section 3.1. Thereafter, Section 3.2 describes our first approach to generate a spatio-temporal dust predictor. It extends LDA using non-parametrically modeled functions represented by GMRFs over estimates of the data under pristine sky conditions. This method, called latent projection functions (LPFs), is then generalized in Section 3.3 to latent signal mapping (LSM). LSM allows for arbitrary mappings instead of linear data projections and thereby reduces noise carried over to the dust predictor. Lastly, Section 3.4 describes an

integration of surface emissivity estimates to the domain of LSM and employs a cross-validation study to demonstrate a superior detection accuracy as compared to other methods.

Having an (approximation) to the dust density available, Chapter 4 develops methods to infer the underlying atmospheric transport process. For this purpose, Section 4.1 provides the reader with basic knowledge concerning fluid dynamics, the Horn and Schunck (HS) approach to flow estimation, variational methods in general and a class of SPDEs that was recently pointed out to be related to the estimation of spatio-temporal correlation. Section 4.2 then elucidates on a probabilistic interpretation of the HS approach, i.e. a BHM in which the unknown variables correspond to the flow field we yield to determine as well as the hyper-parameters of the method. First results are shown for flow fields determined using the INLA technique. In Section 4.3 we analyze how susceptible such flow estimates are to different choices of the dust prediction method and in particular to the noise level the respective predictors come with. A major caveat of the HS approach that also holds for our corresponding BHM is the assumption of a divergence free (incompressible) flow field. Section 4.4 elaborates on how to dismiss this restriction and introduces a BHM for estimation of compressible flow fields. Simulation and case studies are employed to show how this approach clearly reduces errors in the estimated field and yields procedures to forecast dust events or rewind a given dust plume to its source. The Chapter is concluded by Section 4.5, a mathematical analysis of compressible flow and its connection to the material density via the continuity equation.

In Chapter 5 we provide a summarizing conclusion on the work this thesis presents. Concomitantly, we point out a multitude of directions for future research on the topic of probabilistic motion estimation.

2 Graphical Models

Graphical models (GMs) reside at the intersection of two well established mathematical fields. Graph theory provides a framework to express structure primarily from a discrete point of view where objects of interest are either related or not. Probability theory on the other hand has its strengths in the analytical characterization of such relationships. This chapter provides the foundations of both scientific disciplines, introduces graphical models (GMs) and consecutively gives insight into two types of models employed in the context of this thesis as well as the respective algorithmic framework for computational inference.

2.1 Probability Theory

The foundations of modern probability theory, the branch of mathematics that yields to express the intrinsic uncertainty of random phenomena, were laid by the axiomatic basis coined by Andrey Nikolaevich Kolmogorov. These culminate to the core object of the theory, the *probability space*.

Definition 1 (Probability Space)**.** *A probability space $W = (\Omega, \mathcal{F}, P)$ is a triple of a sample space Ω, a σ-algebra (a set of events) $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ and probability measure $P : \mathcal{F} \to [0, 1] \subseteq \mathbb{R}$.*

In an experimental setting, the sample space governs the possible outcomes or observations. Yet, the quantity corresponding to the intuitive likelihood that is intended to be modeled is not directly associated with the sample space. The *probability measure*, which has to fulfil the Kolmogorov axioms and expresses probabilities in terms of a real number, is defined on a set of events $\mathcal{F}$ called σ-algebra.

Definition 2 (σ-Algebra)**.** *A σ-Algebra $\mathcal{F}$ of a set X is a non-empty subset of the power set $\mathcal{P}(X)$ such that the following conditions are fulfilled*

- $X \in \mathcal{F}$
- *$\mathcal{F}$ is closed under complementation: if $A \in \mathcal{F}$ then $X \setminus A \in \mathcal{F}$*
- *$\mathcal{F}$ is closed under countable unions: let $\{A_i \in \mathcal{F}\}_{i \in \mathcal{I}}$ then $\mathcal{F} \ni \cup_{\mathcal{I}} A_i$,*

where $\mathcal{I}$ is a countable index set.

Note that the closure under countable set intersections then follows from the De Morgan's laws. A subtlety of this approach is that it abstracts from the type of events that are considered. In particular this depends upon the cardinality of the sample space, i.e. whether it is countable and thus discrete events are modeled or uncountable infinite as in assigning a probabilities to subsets of a continuous domain like the real numbers. For the sake of simplicity of the document at hand a notation will be adapted that suffices to express both cases in a adequate fashion. In both cases we will term an event x a realization of a *random variable* $\mathcal{X}$ and will do not differentiate between a probability mass function $P(\mathcal{X})$ and a probability density function $p(x)$ by using the generalizing term *probability distribution*.

Vector valued random variables give rise to a characterization of the dependency between different events. Alongside the *marginal distribution* $p(x)$ of a random variable $\mathcal{X}$ one then analyses the *joint distribution* $p(x, y)$ of $\mathcal{X}$ with $\mathcal{Y}$. From the Kolmogorov axioms it then follows that

$$p(x) = \int p(x, y)\ \mathrm{d}y \quad \text{and} \quad \int p(x)\ \mathrm{d}x = 1,$$

where the integral stands for a summation in the discrete case. Subsequently, the *conditional distribution* defined as $p(y|x) = p(x, y)/p(x)$ allows to assess the properties of a *hidden* or *latent* random variable $\mathcal{Y}$ if the outcome x of $\mathcal{X}$ is assumed to be known. With respect to this, a major result of probability theory is expressed by *Bayes' theorem*

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)}, \tag{2.1}$$

which stipulates how the conditional $p(y|x)$ relates to $p(x|y)$. In this context, the former is referred to as the *posterior distribution* of y given x, $p(x|y)$ is the *likelihood* of the observations x and $p(y)$ is the *prior distribution* (or simply prior) the hidden variables are assumed to come from. The denominator $p(x)$, called *normalization constant*, plays a special role within this framework when it comes to actual inference on the latent variables. Although there has been a steady methodological progress the normalization constant typically comes with the intricacy of analytical and/or computational intractability.

Throughout the next sections the concept of *(conditional) independence* plays an important role. Two random variables $\mathcal{X}$ and $\mathcal{Y}$ are called independent (denoted by $\mathcal{X} \perp\!\!\!\perp \mathcal{Y}$) if and only if their joint distribution $p(x, y)$ equals the product of the respective marginal distributions, i.e. $p(x, y) = p(x)p(y)$. From the definition of the conditional distribution it is then easy to see that this implies equality of the conditional to the marginal distribution for both variables and hence that knowledge about the state of one variable does not result in additional knowledge about the state of the other. Yet, even though two variables might not be independent, their dependence might be resolved by a third variable $\mathcal{Z}$. Accordingly, $\mathcal{X}$ and $\mathcal{Y}$ are called *conditionally independent* given $\mathcal{Z}$ (denoted by $\mathcal{X} \perp\!\!\!\perp \mathcal{Y}|\mathcal{Z}$) if and only if the

conditional distributions factorize, i.e. whenever $p(x, y|z) = p(x|z)p(y|z)$. Conditional independence and its important role in graphical models will be elucidated in section 2.3 where the following theorem is applied extensively.

Theorem 1 (Factorization criterion for conditional independence)**.** *Given a distribution $p(x, y, z)$ the two random variables $\mathcal{X}$ and $\mathcal{Y}$ are conditionally independent given $\mathcal{Z}$ if and only if p factorizes into components depending on either x or y but not both, i.e.*

$$x \perp y|z \Leftrightarrow p(x, y, z) = f(x, z)g(y, z) \tag{2.2}$$

for some functions f and g and for all z with $p(z) > 0$.

A detailed explanation on this an related theorems is given by Lauritzen (1996, Chapter 3.2). As a last point, we will make use of the *expectation* $\mathbb{E}$ of a function f under a random variable $\mathcal{X}$,

$$\mathbb{E}_f[\mathcal{X}] = \int f(x)p(x)\ \mathrm{d}x,$$

as well as the mean,

$$\bar{\mathcal{X}} = \mathbb{E}_{\mathrm{id}}[\mathcal{X}] = \int xp(x)\ \mathrm{d}x$$

and the variance,

$$\mathrm{Var}(\mathcal{X}) = \mathbb{E}_{(x-\bar{\mathcal{X}})^2}[\mathcal{X}] = \int (x - \bar{\mathcal{X}})^2 p(x)\ \mathrm{d}x.$$

2.2 Graph Theory

Ever since Leonhard Euler first formulated the problem of the "Seven Bridges of Königsberg" in 1736 *graph theory* has evolved to a powerful mathematical tool that found application to a wide range of scientific disciplines. The main reason for this is the simplicity of the involved mathematical objects as well as their straight forward and intuitive approach to express relationships in a way that is congruent with human perception of the every day mechanics of the world. The following paragraphs introduce graphs as well as some of their properties and respective terminology as a guide through the remainder of this thesis. A graph is an ordered pair of two entities, a set of nodes (vertices) that represent objects and a set of edges each representing a relation among two of these nodes. The following definition is embraced throughout this thesis.

Definition 3 (Graph)**.** *A graph $G = (V, E)$ is a tuple of a set of vertices V and a set of edges $E \subseteq \{(a, b) \in V \times V | a \neq b\}$.*

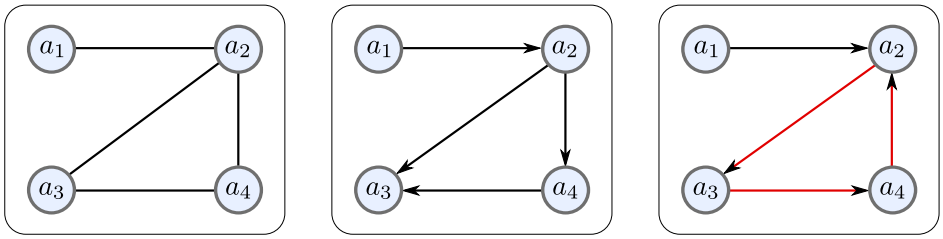


Figure 2.1: Directed and undirected Graphs. Panel (a) depicts an undirected graph with vertices $V = \{a_1, a_2, a_3, a_4\}$ and edge set $E = \{\{a_1, a_2\}, \{a_2, a_3\}, \{a_2, a_4\}, \{a_3, a_4\}\}$. Panels (b) and (c) both depict directed graphs and indicate the direction of an edge by an arrow. While the graph in Panel (c) is cyclic due to the path (a_2, a_3, a_4) the graph in Panel (b) is free of such cycles and called a DAG.

So far, no assumption about the symmetry of these relations is made except for the exclusion of edges from a node to itself. This changes when considering either *directed* or *undirected* graphs. Figure 2.1 displays a typical way of sketching these. In undirected graphs (see Figure 2.1(a)) the edges are unordered pairs, i.e. one has that an edge $e = (a, b)$ equals a set $e = \{a, b\} = \{b, a\}$. The relation this edge encodes for the two nodes is thus understood to be symmetric. Directed graphs take a different approach. They represent edges as ordered pairs $e = (a, b)$ for which the order reflects a direction of the relation. That is, for instance, the causality of two consecutive events. Figure 2.1(b) shows how this idea leads to a graph representation modeling a chain of events that influence each other. The interpretation of an edge in terms of causality reveals an intricacy of these graphs that is pointed out by Figure 2.1(c). In a classical sense causality requires that an event that influences another event has to precede the latter in time. As this sets forth to chains of events it is common practice to only consider *acyclic graphs* as opposed to *cyclic* graphs. The proper definition of these graph types as well as other characterizations of graph properties is greatly facilitated by the following definitions.

Definition 4 (Relations in a Graph)**.** *Let $G = (V, E)$ be a directed or undirected graph, $A \subseteq V$ and $a \in V$. Then*

$$\begin{aligned}
\mathrm{ne}(A) &= \{b \in V \setminus A | \exists a \in A : \{(a, b), (b, a), \{a, b\}\} \cap E \neq \emptyset\} && (2.3)\\
\deg(a) &= |\mathrm{ne}(a)| && (2.4)\\
\mathrm{pa}(A) &= \{b \in V \setminus A | \exists a \in A : (b, a) \in E\} && (2.5)\\
\mathrm{ch}(A) &= \{b \in V \setminus A | \exists a \in A : (a, b) \in E\} && (2.6)\\
\mathrm{de}(A) &= \{b \in V | \exists (c_1, \ldots, c_k) : c_1 \in A, c_k = b, \\
&\quad \forall i \in \{1, \ldots, k-1\}(c_i, c_{i+1}) \in E\}, && (2.7)
\end{aligned}$$

where for $f \in \{\mathrm{ne}, \mathrm{pa}, \mathrm{ch}, \mathrm{de}\}$ $f(a)$ denotes $f(\{a\})$.

These can be thought of as follows. If there is an edge connecting two nodes, these nodes are said to be *neighbors* w.r.t. the given graph. The set of neighbors of a node $a \in V$ or a set of nodes $A \subseteq V$ is denoted by $\mathrm{ne}(a)$ and $\mathrm{ne}(A)$, respectively (see Equation (2.3)). Counting the number of neighbors of a node results in the *degree* $\deg(a)$ of the node. Throughout this thesis an emphasis is put on graphs for which the degree of the nodes is small compared to the total number of nodes in the respective graph. In this case and by embracing the nomenclature popular for general matrices (which will later encode the edges of a graph) a graph is said to be *sparse*. If, on the other hand, the degree of the nodes is relatively high the graph is called *dense*. The respective limiting case, a *complete* graph, is a graph for which each node is connected by an edge to all of the other nodes of the graph.

In directed graphs the edges are are not symmetric and hence encode more than just neighborhood in the graph. If there is a directed edge from a node a to a node b then a is said to be a *parent* of node b. This is set forth by the definition of the set of parents $\mathrm{pa}(A)$ of a subset $A \subseteq V$ of the nodes (see Equation (2.5)), which accumulates all parents of the nodes in A. Conversely, b is called a *child* of a if graph contains a directed edge from a to b. The set of all children of a subset $A \subseteq V$ is defined in Equation (2.6).

The property of neighborship is set forth in a transitive fashion by considering *paths* along the edges of a graph. That is, a path is a sequence $(a_1, \ldots, a_k)$ of distinct nodes of the graph such that for each $i \in \{1, \ldots, k-1\}$ the nodes a_i and a_{i+1} are neighbors. The length of this path is the number of contained edges $k-1$. Two nodes a and b are then said to be *connected* with respect to a given graph if and only if there exists at least one path between a and b. If all edges on a path from a to b are directed to b the path is called *directed* as well. Using paths a precise definition of acyclic graphs is advisable. The descendants of a node a (see Equation (2.7)) constitute a set of nodes for which a directed path exists from a to the respective node. Now, if a is an element of its own descendants, it is easy to see that there exists a directed path from node a to itself. This *directed cycle* is in dissent with a potential cause-and-effect interpretation of the relations modeled by the graph. A widespread strategy is thus to only consider DAGs, which are the basis of Bayesian networks to be discussed in Section 2.3.2. These preclude the occurrence of directed cycles.

In general, paths play an integral role in the analysis of probabilistic interpretations of graphs, i.e. probabilistic graphical models (PGMs). They allow to define so called *separation properties* that characterize the conditional independence assumptions of a probability distribution associated with the respective graph. At the same time, paths are a very intuitive mode of thinking in terms of network connectivity and hence form the basis of a wide range of algorithms based on graph structures, e.g. by solving combinatorial problems with graph cuts (Boykov et al., 2001). The following sections go into more detail on PGMs and their practical implications.

2.3 Probabilistic Graphical Models

With the fundamentals of probability and graph theory at hand the definition of probabilistic graphical models (PGMs) is straight forward. A PGM M is a pair $M = (\mathcal{X}, G)$ of a random vector $\mathcal{X}$ and a graph $G = (V, E)$. Each of the nodes of the latter is associated with a component of the random vector via a bijective mapping. This mapping is employed to index the components of the random vector by the nodes of the graph, i.e., the component of $\mathcal{X}$ that is represented by a node $a \in V$ is written as $\mathcal{X}_a$ with realization x_a. The indexed component of $\mathcal{X}$ can again be a random vector or a univariate random variable. In may cases an indexing by a subset $\mathcal{I}$ of the natural numbers eases up the notation. A node is then referred to as node $i \in V = \mathcal{I} \subseteq \mathbb{N}$ with corresponding random component $\mathcal{X}_i$. The involved part of the connection between the graph and the random vector concerns the actual distribution $p(x)$ and the graph structure. This is apparent from the following definition.

Definition 5 (Probabilistic Graphical Model)**.** *A PGM is a pair $(\mathcal{X}, G)$ of a graph $G = (V, E)$ and a random vector $\mathcal{X}$. The graph defines a ternary relation $(\cdot \perp\!\!\!\perp_G \cdot \,|\, \cdot)$ such that for all $S \subset V$ and any disjoint sets $A, B \subset V \setminus S$ it holds that*

$$A \perp\!\!\!\perp_G B|S \Rightarrow X_A \perp\!\!\!\perp X_B|X_S. \tag{2.8}$$

The intuition behind this Definition 5 is that the graph allows to deduce the conditional independence structure of the random vector from its edges and vertices. At the same time the factorization criterion (Theorem 1) then paves the way to deduce information about the analytic structure of the distribution p. For instance, if a component X_S is given by an observation and inference on X_A and X_B is to be performed, G might stipulate that the respective posteriors can be assessed independently, e.g., in parallel. The definition of the relation $\perp\!\!\!\perp_G$, however, intrinsically depends strongly on the type of graph that is considered.

Undirected graphical models are examined in more detail in section 2.3.1. For these models the rather simple separation property is sufficient to define the relation $\perp\!\!\!\perp_G$. The independence statements that follow from this property are commonly referred to as *Markov properties*, which lead to the widespread nomenclature of *Markov networks* and *Markov random fields (MRFs)*.

For directed graphs the more involved formulation of *d-separation* coined by Pearl (1988) is required. A comprehensive overview on the topic of PGMs is also given by Koller and Friedman (2009), who additionally provide a perspective of mapping assertions about graphs to distributions and vice versa. A short introduction to these models, also known as Bayesian Networks, is given in section 2.3.2.

2.3.1 Undirected Graphical Models (Markov Networks)

In undirected GMs, also known as Markov random fields (MRFs) or Markov networks the definition of the relation $\perp\!\!\!\perp_G$ is based on the following separation property.

Definition 6 (Separation)**.** *Let $G = (V, E)$ be an undirected graph and let $S, A, B \subset V$ be subsets of the vertices with $A \cap B = \emptyset$. Then S is said to separate $A, B \subset V \setminus S$ if in the subgraph $G_{V \setminus S}$ there exists no path between any node in A and any node in B. The separation property is denoted by the relation $(\cdot \perp_G \cdot | \cdot)$.*

The definition of an undirected GM is then as follows.

Definition 7 (Undirected Graphical Model)**.** *An undirected graphical model M is a graphical model $M = (\mathcal{X}, G)$ with relation $\perp\!\!\!\perp_G$ defined via the separation property, i.e. $\perp\!\!\!\perp_G := \perp_G$.*

Here it is apparent that, by definition, undirected GMs imply conditional independence by the lack of particular paths between the sets of nodes that represent the variables under consideration. This feature is the first in a list of so called *Markov properties*:

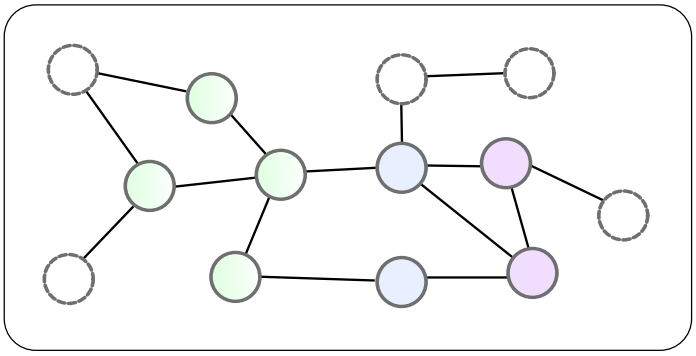
Definition 8 (Markov Properties)**.** *Let $G = (V, E)$ be an undirected graph and X be a random vector with index set $\mathcal{V}$. We then define that X has*

(G) the global Markov property with respect to G, if for each triple (A, B, S) of disjoint subsets of V such that S separates A and B in G the conditional independence $X_A \perp\!\!\!\perp X_B | X_S$ holds.

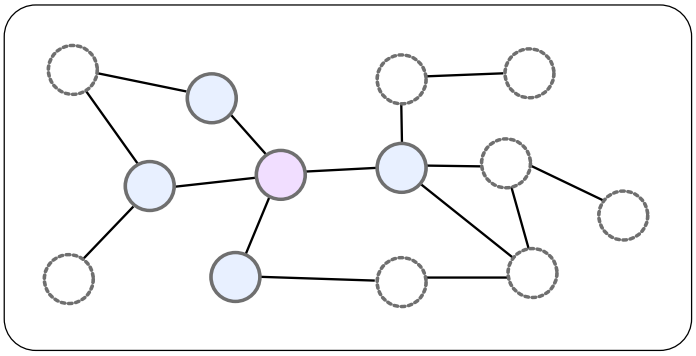
(L) the local Markov property with respect to G, if for all nodes $i \in V$ the conditional independence $X_i \perp\!\!\!\perp X_{V \setminus \{\mathrm{ne}(i), i\}} | X_{\mathrm{ne}(i)}$ holds.

(P) the pairwise Markov property with respect to G, if for all pairs of non-adjacent nodes i and j, the conditional independence $X_i \perp\!\!\!\perp X_j | X_{V \setminus \{i,j\}}$ holds.

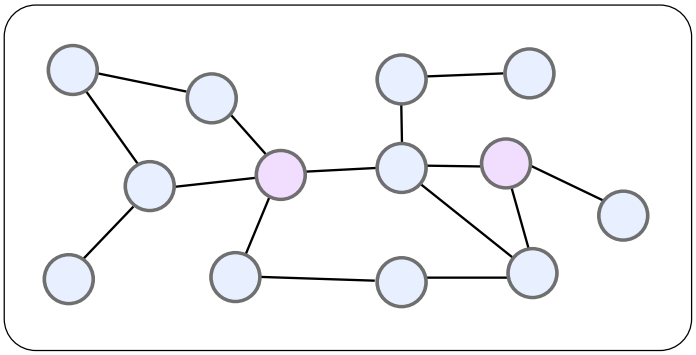
Figure 2.2 provides an intuition on the character of these properties. When comparing the respective Panels (a) and (b) it is easy to see that the local Markov property is a special case of the global property as the neighbors (blue) of the red node in (b) separate the red node from the other nodes. The blue nodes in this Panel act as what is commonly referred to as a *Markov blanket* for the red node. Likewise, the pairwise Markov property is a special case of the local Markov property. Since the two nodes in (c) are not adjacent, each of their sets of neighbors can act as a Markov blanket that is a subset of the blue nodes, respectively. Theorem 2 summarizes these observations.



(a)



(b)



(c)

Figure 2.2: Illustration of the Markov properties, adapted from Rue and Held (2005). Panel (a) depicts the global Markov property. The blue nodes separate the red and green nodes, which are thus independent given the former. The local Markov property, shown in Panel (b), stipulates that conditional on its neighbors (blue) the red node is independent from all other nodes of the graph (white). The two red nodes in Panel 2.2(c) are independent given the blue nodes since every path that connects them passes a blue node. Hence, the pairwise Markov property holds.

Theorem 2 (Markov Property Implications)**.** *For an arbitrary random vector $\mathcal{X}$ the global Markov property (G) implies the local Markov property (L) and the local Markov property implies the pairwise Markov property (P), i.e.;*

$$(G) \Rightarrow (L) \Rightarrow (P). \tag{2.9}$$

Moreover, a under a mild restriction it is possible to show that these implications are circular, i.e., that the pairwise Markov property implies the global Markov property. It suffices to require that the distribution p ensures that

$$X \perp\!\!\!\perp Y|Z \wedge X \perp\!\!\!\perp Z|Y \Rightarrow X \perp\!\!\!\perp (Y, Z)$$

holds. As shown by Pearl and Paz (1985) this can be achieved by restricting the model to strictly positive distributions:

Theorem 3 (Equivalence of Markov Properties)**.** *For a random vector with a strict positive distribution, i.e., $p(x) > 0$, the pairwise Markov property implies the global Markov property, i.e.: (P) ⇒(G).*

For undirected Graphical models with a strictly positive distribution of $\mathcal{X}$ the Markov properties are thus equivalent. An example for such a class of models are Gaussian Markov random fields (GMRFs), which are introduced in section 2.6.

2.3.2 Directed Graphical Models (Bayesian Networks)

Bayesian Networks or directed graphical models (DGMs) lend themselves towards expressing causal relationships among the incorporated randoms variable and in particular between a node and its parents. Bayes theorem, on the other hand, allows to characterize the joint distribution of two random variables as proportional to the product of a conditional and a marginal. It does not come as a surprise that the most common definition of a DGM builds upon this similarity.

Definition 9 (Directed Graphical Model)**.** *A directed graphical model M is a pair $M = (\mathcal{X}, G)$ of a random vector $\mathcal{X}$ and a directed acyclic graph $G = (V, E)$ such that the joint distribution $p(x)$ of $\mathcal{X}$ factorizes into a product of conditional distributions of the form*

$$p(x) = \prod_{a \in V} p(x_a | x_{\mathrm{pa}(a)}). \tag{2.10}$$

As opposed to undirected GMs this definition does not explicitly state a ternary relation $\perp\!\!\!\perp_G$ implying conditional independence. However, using the notion of the *d-separation*, it is possible to show that the class of probabilistic models defined by Equation (2.10) is characterized by such a relation as well. Throughout this thesis DGMs appear in the form of Bayesian hierarchical models (BHMs), which

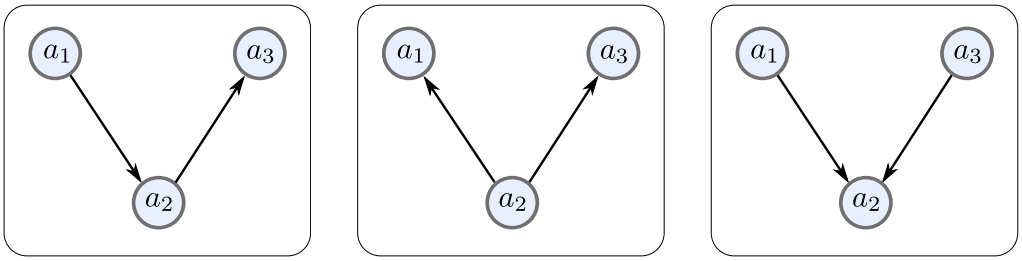


Figure 2.3: Path types in DGMs, adapted from Bishop (2006). Node a_2 is said so be (a) head-to-tail, (b) tail-to-tail or (c) head-to-head with respect to the path (a_1, a_2, a_3).

will be introduced in section 2.4. As these models are of rather simple nature in terms of their graph structure only the main conditional independence Theorem that incorporates d-separation is stated here. It allows to read the conditional independence assumptions to be read directly from the graph and thus without further inspection of probabilistic properties of the distribution.

Theorem 4. *In a directed graph $G = (V, E)$ the set $S \subset V$ d-separates the disjoint subsets $A, B \subset V \setminus S$ (denoted by $A \perp_G B|S$) if S blocks all paths between A and B. A path between a node $a \in A$ and a node $b \in B$ is defined to be blocked if it includes a node s such that either*

i) $s \in S$ and s is head-to-tail or tail-to-head node in this path

ii) $(\{s \cup \mathrm{ne}(s)\}) \cap S = \emptyset$ and s is head-to-head node in this path.

For a directed graphical model (GM) $M = (\mathcal{X}, G)$ the relation $\perp_G$ implies conditional independence, i.e.

$$A \perp_G B|S \Rightarrow \mathcal{X}_A \perp\!\!\!\perp \mathcal{X}_B|\mathcal{X}_S.$$

More details on the theorem and a proof are given by Pearl (1988). Figure 2.3 provides an intuition on the terminology of head-to-tail, tail-to-tail and head-to-tail blocking of a path.

2.4 Bayesian Hierarchical Models

With the increase of amount and complexity of collected data during the last decades probabilistic models have become common that structurally reflect the generative nature of the observations at hand. As opposed to building a marginal model by defining a data likelihood for which parameters are inferred, hierarchical models (HMs) yield a clear separation between the involved stochastic components (Cressie and

Wikle, 2011). Following the terminology of Berliner (1996), the general idea is to distinguish between a bottom layer comprising a *process model* and top layer that defines the *data model* conditional on the hidden process. On the one hand, this allows to think of the observed data as a realization from a distribution reflecting the intrinsic uncertainty of a measurement process. On the other hand, it is assumed that the outcome of the measurement depends on a realization of a "true" (hidden or latent) process that has a distribution on its own and thereby represents the uncertainty about assumptions on the underlying phenomenon the data is assumed to be intrinsically generated by. Typically, these distributions are specified such as to depend on further parameters that reside on an even lower level of the hierarchy and are unknown as well. Models of this kind are commonly referred to as BHMs if the distribution of this *parameter layer* is modeled jointly with the distribution of the other unknown variables. An alternative approach is to assume that the parameters are fixed and to estimate them using the data. This is commonly referred to as a empirical hierarchical model (EHM) and inference on the process layer proceeds by substitution of the fixed parameters into the process and data layers.

Due to the definition of conditional probabilities it is straightforward to put the joint distribution of a BHM into an analytic form that reflects the aforementioned causal structure. Let $\mathbf{y}$ be the random random vector of observed data, $\mathbf{x}$ the hidden process and Θ the respective random parameters. We then have that

$$p(\mathbf{y}, \mathbf{x}, \boldsymbol{\Theta}) = p(\mathbf{y}|\mathbf{x}, \boldsymbol{\Theta})p(\mathbf{x}|\boldsymbol{\Theta})p(\boldsymbol{\Theta}). \tag{2.11}$$

A graphical representation of this BHM is given in Figure 2.4. It also also holds a second graph that results from assuming independent sets of parameters Θ_D and Θ_P for the data and process model, a situation not uncommon among BHMs applications. Bayes' theorem now gives the posterior distribution of the process model and the parameters. Hence, by expressing the denominator $p(\mathbf{y})$ in terms of an integrated joint distribution one obtains

$$\pi(\mathbf{x}, \boldsymbol{\Theta}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x}, \boldsymbol{\Theta})p(\mathbf{x}|\boldsymbol{\Theta})p(\boldsymbol{\Theta})}{\iint p(\mathbf{y}|\mathbf{x}, \boldsymbol{\Theta})p(\mathbf{x}|\boldsymbol{\Theta})p(\boldsymbol{\Theta})\mathrm{d}\boldsymbol{\Theta}\mathrm{d}\mathbf{x}}. \tag{2.12}$$

The denominator of Equation (2.12), i.e. the normalization constant that ensures that the total probability equals 1, also illustrates the intricacy anticipated in Section 2.1. When no explicit marginal distribution for the data can be assumed (which could also be seen as a dissent to the BHM paradigm itself), analytical or numerical integration is inevitable. For special cases where the analytic integration is feasible or the dimensionality of the latent variables and the parameters is small and therefore methods like quadrature can be applied this does not pose a problem. However, especially in applications with spatio-temporal models the dimensions can easily get on the order of thousands and more sophisticated methods are required.

A class of computational approaches that lead to respective breakthroughs during the last twenty years are Monte Carlo methods like Markov chain Monte Carlo (MCMC)

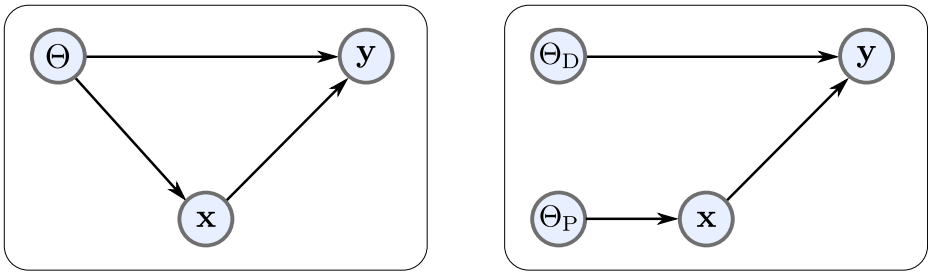


Figure 2.4: Graphical representation of a BHM, adapted from Cressie and Wikle (2011). (a) Observations Y depend on the parameters Θ as well as the hidden process model X. The latter only depends on the parameters. (b) The process model and the observation model are assumed to have independent sets of parameters Θ_P and Θ_D which are therefore not connected by an edge.

and importance sampling (IS) (e.g. Robert and Casella (2004)). Here, the strong law of large numbers is employed to construct approximate expectation values

$$\mathbb{E}g(\mathcal{X}) \approx (1/K) \sum_{k=1}^{K} g(\mathcal{X}_k), \tag{2.13}$$

and to characterize analytically intractable distributions via samples $\mathcal{X}_1 \dots \mathcal{X}_K$ (where the approximation improves with increasing K). For example, a well-considered sampling scheme and $g \equiv p$ leads to an estimator of the aforementioned integral constituting the normalization constant.

Although Monte Carlo methods for HMs have encountered a steep ascendance since their first application to image analysis by Geman and Geman (1984) they are not without detriments. *Simulating* from a joint distribution commonly requires extensive analysis on the usually incorporated sub-sampling steps, an assessment of the suitability of the drawn samples (e.g. their autocorrelation along the sampling process) as well as expensive and extensive computation time. However, recent developments of deterministic approximation methods have paved the way for a special case of this issue. When the objects of interest are the marginal distributions of the involved process components and the parameters, INLAs can serve as a mean for fast and almost exact inference. As this approach is a core component of the experiments conducted for this thesis it is explained in greater detail in Section 2.7.

2.5 Generalized Linear Models and the Exponential Family

The notation of GLMs goes back to Nelder and Wedderburn (1972) and their successful attempt to unify different methods of regression analysis with non-normal responses (observations). In the classical setting for deterministic covariates and ungrouped normal responses it is assumed that the relation

$$y_i = \mathbf{z}_i^T \boldsymbol{\beta} + \epsilon_i, \qquad i = 1, \ldots, n, \qquad \epsilon_i \sim \mathcal{N}(0, \sigma^2) \tag{2.14}$$

holds, where the design vector $\mathbf{z}_i = (1, \mathbf{x}_i)$ is an appropriate function of the covariate vector $\mathbf{x}_i$ and $\boldsymbol{\beta}$ is a vector of unknown parameters. Rewriting the model such that the observations are stated as independent and identically distributed given the expectation $\mu_i = \mathbb{E}(\mathbf{y}_i) = \mathbf{z}_i^T \boldsymbol{\beta}_i$ with

$$y_i \sim \mathcal{N}(\mu_i, \sigma^2),$$

leads to a conditional understanding of the former model. Hence, if covariates are stochastic, the pairs $(\mathbf{x}_i, y_i)$ are assumed to be iid and Equation (2.5) gives the conditional density of the conditionally independent y_i given the $\mathbf{x}_i$. Relaxing these preceding assumptions w.r.t. the type of expectation and conditional distributions then naturally extends the covered family of models and we employ the following definition of GLMs stated by Fahrmeir and Tutz (2001).

Definition 10 (Generalized linear model)**.** *A GLM extends the linear model 2.14 by making the following assumptions:*

1. Distribution assumption: Given $\mathbf{x}_i$, the y_i are (conditionally) independent, and the (conditional) distribution of y_i belongs to a *simple exponential family* with (conditional) expectation
$$\mathbb{E}(y_i|\mathbf{x}_i) = \mu_i \tag{2.15}$$
and , possibly, a common scale parameter ϕ, not depending on i.

2. Structural assumption: The expectation μ_i is related to the linear predictor $\eta_i = \mathbf{z}_i^T \boldsymbol{\beta}$ by
$$\mu_i = h(\eta_i) = h(\mathbf{z}_i^T \boldsymbol{\beta}) \qquad \text{resp.,} \quad \eta_i = g(\mu_i), \tag{2.16}$$
where

   h is a known one-to-one, sufficiently smooth response function,

   g is the link function, i.e., the inverse of h,

   $\boldsymbol{\beta}$ is a vector of unknown parameters of dimension p, and

   $\mathbf{z}_i$ is a design vector of dimension p, which is determined as an appropriate function $\mathbf{z}_i = z(\mathbf{x}_i)$ of the covariates.

This definition requires acquaintance with the notion of a simple exponential family of probability distributions (see also Fahrmeir and Tutz (2001), Appendix A1).

Definition 11 (Simple Exponential Family)**.** *The distribution p of a random variable y belongs to a simple exponential family if its discrete or continuous density with respect to a σ-finite[1] measure*

$$p(y|\theta,\phi) = \exp\Big[\frac{y\theta - b(\theta)}{\phi} + c(y,\phi)\Big], \tag{2.17}$$

with $c(y,\phi) \geq 0$ and measurable. The functions b and c are specific to the type of the exponential family for which θ is called natural parameter and ϕ is a nuisance or dispersion parameter.

It should be noted that this definition restricts the interaction term $y\theta$ to a product of the observations and the parameters, which is not the case in general exponential families where a function $T(y)$ substitutes y. In the vector valued case, the mentioned product translates to an inner product of the observation and parameter space. An alternative nomenclature is to refer to the restricted class of models as *natural* exponential families (see Robert (2007)). Distributions of this type have properties that make them particularly appealing for statistical analysis.

- When considering a set of iid samples $y_1, \ldots, y_k \sim f(y,\theta)$ there exists a *sufficient statistic* $S(y_1, \ldots y_k)$ of constant dimension. Here, $S(\cdot)$ is called a sufficient statistic if the distribution of $y_1, \ldots y_k$ conditional upon $S(y_1, \ldots y_k)$ does not depend on θ. This means that the statistic contains the whole information brought by $y_1, \ldots y_k$ about θ. In case of natural exponential families, a sufficient statistic is the mean
$$\bar{y} = 1/k \sum_{j=1}^{k} y_j,$$
which implies that inference on θ only depends on a small summary of a possibly large data set.

- Members of the exponential families are guaranteed to have *conjugate priors*. A conjugate prior of $p(y|\theta)$ is a prior distribution $p(\theta)$ from a family $\mathcal{F}$ of the parameters such that under the application of Bayes' theorem the posterior $p(\theta|y)$ also belongs to $\mathcal{F}$. Hence, if the prior has an analytically or computationally convenient form this carries forth to the posterior.

- The first and second order moments of the distribution $p(y|\theta,\phi)$ follow directly from the choice of the function $b(\cdot)$. This also holds if y is a random vector $\mathbf{y} = y$ and $\Theta = \theta$ is vector valued as well, i.e. when considering $p(\mathbf{y}|\Theta,\phi)$. To be precise, if $\tilde{\theta}$ is the set of all θ satisfying $0 < \int p \, \mathrm{d}\mathbf{y} < \infty$, then $\tilde{\theta}$ is convex

[1] A measure μ is said to be σ-finite if there is a countable set $\{A_i \in \mathcal{F}\}_{i\in\mathcal{I}}$ of σ-finite A_i's ($\mu(A_i) < \infty$ for all $i \in \mathcal{I}$) such that the sample space Ω is a union of these sets ($\Omega = \cup_{\mathcal{I}} A_i$).

Table 2.1: Simple exponential families of the form (a) with analytic components (b)

(a) $$f(y|\theta,\phi,\omega) = \exp\left[\frac{y\theta-\beta(\theta)}{\phi}\omega + c(y,\phi,\omega)\right]$$

(b) Components

Distribution		$\theta(\mu)$	$\beta(\theta)$	ϕ	c
Normal	$\mathcal{N}(\mu,\sigma^2)$	μ	$\theta^2/2$	σ^2	$y^2/2\sigma^2$
Bernoulli	$\mathcal{B}(1,\pi)$	$\log(\pi/(1-\pi))$	$\log(1+\exp(\theta))$	1	0
Gamma	$\mathcal{G}(\mu,\nu)$	$-1/\mu$	$-\log(-\theta)$	ν^{-1}	$\log(\frac{y^{\mu-1}}{\Gamma(\mu)})$

and in the interior, assumed to be nonvoid, all derivatives of b and all moments of $\mathbf{y}$ exist. In particular

$$\begin{aligned}\mathbb{E}(\mathbf{y}|\theta) &= \mu(\theta) = \frac{\partial b}{\partial\theta}\\ \mathrm{Cov}(\mathbf{y}|\boldsymbol{\theta}) &= \Sigma(\theta) = \frac{\partial^2 b}{\partial\theta\partial\theta'}\end{aligned}$$

The latter equations play an important role with respect to the specification of GLMs. It has to be taken into account that a particular mean structure given by $b(\cdot)$ implies a certain covariance structure and that the natural parameter is a uniquely determined function $\Theta(\mu)$ of the mean given a particular exponential family. This influences the choice of the response or link function. For each exponential family there exists a natural or canonical link function that relates the natural parameter directly to the linear predictor:

$$\theta(\mu) = \eta = \mathbf{z}^T\boldsymbol{\beta},$$

and hence as $\eta = g(\mu)$ (Eq. 2.16):

$$\theta(\mu) \equiv g(\mu).$$

From this, models with convenient mathematical and statistical properties evolve. Prominent examples are the normal, Bernoulli and Gamma distributions, for which the respective link functions can be determined from Table 2.1.

2.6 Gaussian Markov Random Fields

This section has introduced a major building block of the models employed throughout this thesis. Its main constituent is the *Gaussian* or *normal* distribution

$$\mathcal{N}(x \in \mathbb{R}; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{(x-\mu)^2}{2\sigma^2}} \tag{2.18}$$

with mean μ and variance σ^2. Among other favorable analytic and computational properties that set forth to multiple dimension a main motivation to assume normally distributed variables is the central limit theorem. It states that the arithmetic mean of a sufficiently large number of independent random variables with well-defined expected value and variance will be approximately normally distributed. In applied contexts where the actual distribution of variables are rarely known with sufficient certainty, this is an appealing property to maintain a probabilistic foundation for further analysis. The normal distribution is easily set forth to the multivariate case of n dimensions by the definition

$$\mathcal{N}(\mathbf{x} \in \mathbb{R}^n; \boldsymbol{\mu}, \Sigma^{-1}) = (2\pi)^{-n/2} |\Sigma|^{-1/2} \exp\Big(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}-\boldsymbol{\mu}) \Big), \tag{2.19}$$

where $\boldsymbol{\mu}$ is the mean vector and Σ is the so called covariance matrix with determinant $|\Sigma|$. As to define a proper probability distribution the $(n \times n)$ matrix Σ is required to be *positive definite*. This is the case if and only if

$$\mathbf{x}^T \Sigma \mathbf{x} > 0 \qquad \forall\, \mathbf{x} \neq \mathbf{0}$$

and for convenience denoted by $\Sigma > 0$. In the following we will consider the case where Σ is also symmetric, i.e. $\Sigma^T = \Sigma$, and hence symmetric positive definite (SPD).

It turns out that the inverse $Q = \Sigma^{-1}$, called *precision matrix*, plays a key role when considering multivariate normals in context of probabilistic graphical models. From Equation (2.19) it is easy to see that $\mathbb{E}(x_i) = \mu_i$ as well as $\mathrm{Cov}(x_i, x_j) = \Sigma_{ij}$. The covariance matrix thus directly translates into a binary *marginal independence structure* of the distribution :

$$x_i \perp x_j \Leftrightarrow \Sigma_{ij} = 0.$$

The precision matrix, on the other hand, can be shown to reveal another kind of insight into the structure of the distribution:

Theorem 5. *Let* $\mathbf{x}$ *be normal distributed with mean* $\boldsymbol{\mu}$ *and precision matrix* $Q > 0$. *Then for* $i \neq j$,

$$x_i \perp x_j | \mathbf{x}_{-ij} \Leftrightarrow Q_{ij} = 0.$$

The nonzero pattern of the precision matrix thus determines the *conditional independence structure* of the distribution. Finally, as elaborated in Section 2.3.1 this structure is exactly what is needed to formulate an undirected PGM, which gives rise to the following definition:

Definition 12 (Gaussian Markov Random Field (GMRF)). *A random vector* $\mathbf{x} = (x_1, \dots, x_n)^T \in \mathbb{R}^n$ *is called GMRF with respect to a labelled graph* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ *with mean* $\boldsymbol{\mu}$ *and precision matrix* $Q > 0$ *if and only if its density has the form*

$$p(\mathbf{x}) = (2\pi)^{-n/2}|Q|^{1/2} \exp\Big(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T Q(\mathbf{x} - \boldsymbol{\mu})\Big) \tag{2.20}$$

and

$$Q_{ij} \neq 0 \;\Leftrightarrow \{i, j\} \in \mathcal{E} \qquad \forall\, i \neq j.$$

If Q is a completely dense matrix then $\mathcal{G}$ is fully connected. Any normal distribution with SPD covariance matrix is therefore also a GMRF and vice versa. Most importantly, however, is the consequence this has on the Markov properties (see section 8) of the distribution:

Theorem 6 (Markov properties of GMRFs). *Let* $\mathbf{x}$ *be a GMRF with respect to* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ *and distribution* P*. Then the pairwise, the local and the global Markov property are equivalent.*

Proof. Since $p(\mathbf{x}) \propto \exp(\cdot) > 0$ Theorem 3 is applicable. □

From this theorem several powerful results emerge in terms of inference in GMRFs, in particular if sparse precision matrices are considered. These results are best described in an alternative analytic form of the distribution's density, the canonical parameterization.

Definition 13 (Canonical parameterization). *A GMRF* $\mathbf{x}$ *with respect to a graph* $\mathcal{G}$ *with canonical parameters* $\mathbf{b}$ *and* $Q > 0$ *has the density*

$$\mathcal{N}_{\mathrm{C}}(\mathbf{x} \in \mathbb{R}^n; \mathbf{b}, Q) \propto \exp\Big(-\frac{1}{2}\mathbf{x}^T Q\mathbf{x} + \mathbf{b}^T\mathbf{x}\Big). \tag{2.21}$$

Firstly, the mean μ of this distribution is given by $\boldsymbol{\mu} = Q^{-1}\mathbf{b}$ such that one has $\mathcal{N}(\boldsymbol{\mu}, \Sigma^{-1}) = \mathcal{N}_{\mathrm{C}}(Q\boldsymbol{\mu}, Q)$. Switching between representation thus boils down to solving a (sparse) linear system or a matrix vector multiplication. Secondly, the conditional distribution of a subset $A = \mathcal{V} \setminus B$ (A and B being nonempty) of the considered variables is given by the conditional mean

$$\boldsymbol{\mu}_{A|B} = \boldsymbol{\mu}_A - Q_{AA}^{-1}Q_{AB}(\mathbf{x}_B - \boldsymbol{\mu}_B) \tag{2.22}$$

and precision

$$Q_{A|B} = Q_{AA},$$

where without loss of generality a reordering of the variables is considered such that

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_A \\ \boldsymbol{\mu}_B \end{pmatrix} \qquad \text{and} \qquad Q = \begin{pmatrix} Q_{AA} & Q_{AB} \\ Q_{BA} & Q_{BB} \end{pmatrix}.$$

One can see that the conditional dependency structure and hence the structure of a subgraph $\mathcal{G}_A$ representing $\mathbf{x}_A$ is inherited directly from Q, i.e. no computation is needed and $\mathcal{G}_A$ is constructed by simply removing the variables B as well as respective edges. From Equation (2.22) another (computational) advantage arises for sparse graphs. Since Q_{ij} is zero for non neighboring nodes in the graph the conditional mean only depends on values of $\boldsymbol{\mu}$ and Q in A and the immediate neighborhood $\text{ne}(A)$. For small sets A with few neighbors this reduces the necessary computations drastically.

Finally, these favorable properties are even more reflected and complemented by the canonical representation, as can be deduced from the following lemmata:

Lemma 1. *Let* $\mathbf{x} \sim \mathcal{N}_{\text{C}}(\mathbf{b}, Q)$*, then*

$$\mathbf{x}_A | \mathbf{x}_B \sim \mathcal{N}_{\text{C}}(\mathbf{b}_A - Q_{AB}\mathbf{x}_B, Q_{AA})$$

Lemma 2. *Let* $\mathbf{x} \sim \mathcal{N}_{\text{C}}(\mathbf{b}, Q)$ *and* $\mathbf{y}|\mathbf{x} \sim \mathcal{N}(\mathbf{x}, P^{-1})$*, then*

$$\mathbf{x}|\mathbf{y} \sim \mathcal{N}_{\text{C}}(\mathbf{b} + P\mathbf{y}, Q + P)$$

Conditioning a subset of variables from a joint distribution (first lemma) and inference on the posterior with respect to observations $\mathbf{y}$ (second lemma) are hereby both readily available by matrix-vector products. In iterative procedures a successive update of the canonical parameterization is thus very efficient and the computation of the actual mean by solving $Q\boldsymbol{\mu} = \mathbf{b}$ is only performed when actually need, i.e. at the end of the procedure.

Note that also in both cases the sparsity of the precision matrices imply significant reductions in the computational efforts of updating the conditional parameters. For instance, assume that $\mathbf{x}$ is to be conditioned on only one x_i with few connections to k other x_js with respect to the underlying graph. $Q_{AB} = Q_{i,-i}$ in Lemma 1 then reduces to a vector with only k non-zero elements and $\mathbf{x}_B = x_i$ is a scalar. Hence, the canonical parameter $\mathbf{b}_{-i}$ is only updated in k positions. This is one of the key components that render integrated nested Laplace approximations (INLAs) so computationally efficient. More detail on this is provided in the next section.

2.7 Integrated Nested Laplace Approximations

With the upturning popularity of Bayesian models during the last three decades a need for general-purpose inference methods arose. As discussed in section 2.4, MCMC methods provide a main requirement for this generality: they are extremely flexible and are applicable to a wide range of data types and models. Until recently, a main problem remained. Across many scientific fields increasingly sophisticated measurement techniques and the general decrease of costs for data collection led to

complex and computationally highly demanding models. In this case, e.g. spatio-temporal data with a latent Gaussian model on top, MCMC methods typically need extensive run times for sufficiently accurate posterior estimation. For details on this see for instance Rue et al. (2009) and references therein.

2.7.1 Feasible Models

The integrated nested Laplace approximation technique introduced by Rue et al. (2009) provides an extremely powerful approach to the aforementioned computational challenge. It is applicable to the class of structured additive regression models with a latent GMRF and hyper-parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ at the observation and latent layer, respectively. The corresponding BHM has three stages and is structured as follows. The hyper-parameter distribution

$$(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \quad \sim \quad p(\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)) \tag{2.23}$$

is assumed to be low dimensional. First reports stated proper results and reasonable computation times for $K \leq 6$ dimensions. Depending on the structure, even $K \leq 11$ and perhaps beyond can still be dealt with (Finn Lindgren, pers. comm.). The latent variables follow the distribution

$$\mathbf{X} \quad \sim \quad p(\mathbf{x}|\boldsymbol{\theta}_2) = \mathcal{N}(0, Q(\boldsymbol{\theta}_2)), \tag{2.24}$$

i.e. a zero mean GMRF with precision matrix Q parameterized by $\boldsymbol{\theta}_2$. Now, given the parameters $\boldsymbol{\theta}_1$, the observation variables are assumed to be conditionally independent given the latent GMRF:

$$Y_i \quad \overset{iid}{\sim} \quad p(y_i|x_i, \boldsymbol{\theta}_1), \quad i = 1, \ldots, n. \tag{2.25}$$

This way the GMRF provides a flexible mean of modeling different kinds of dependencies between the latent variables of the model that implicitly steer the dependencies among observed data. The observation stage is assumed to be a GLM and the (structured additive) linear predictor is stated as

$$\eta_i = \gamma + \sum_{k=1}^{K} \beta_k z_{ki} + \sum_{l=1}^{L} f_l(u_{li}) + \xi_i, \tag{2.26}$$

where γ denotes a global scalar intercept, the vector $\mathbf{z} = (z_1, \ldots, z_K)$ contains covariates entering the response variable linearly via the coefficients $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)$, $\mathbf{f} = \{f_1(\cdot), \ldots, f_L(\cdot)\}$ is a set of unknown functions of a second covariate vector $\mathbf{u} = (u_1, \ldots, u_L)$ and the ξ_i are unstructured terms. It should be noted that since the functions f_l are modeled by Gaussian variables representing the respective function evaluation this analytic form is equivalent to the general GLM form introduced in section 2.5.

2.7.2 Nested Approximations and Integration

With this model at hand, the major achievement of the INLA technique is efficient inference on the marginal distributions of the unknown variables. In both cases, hyper-parameters and latent field variables, the marginal posteriors are represented by integration over the other hyper-parameter space

$$p(x_i|\mathbf{y}) = \int p(x_i|\mathbf{\Theta},\mathbf{y})p(\mathbf{\Theta}|\mathbf{y})\,\mathrm{d}\mathbf{\Theta} \tag{2.27}$$

$$p(\Theta_j|\mathbf{y}) = \int p(\mathbf{\Theta}|\mathbf{y})\,\mathrm{d}\mathbf{\Theta}_{-j}, \tag{2.28}$$

which explains the respective limitation of INLA to models with low dimensional $\mathbf{\Theta}$. Performing the actual integration requires to evaluation of the full hyper-parameter conditional $p(\mathbf{\Theta}|\mathbf{y})$ as well as the latent marginal posteriors $p(x_i|\mathbf{\Theta},\mathbf{y})$, both of which are not readily available. Instead, Rue et al. (2009) propose the application of the *Laplace approximation* developed by Tierney and Kadane (1986) in both cases and thus in a with respect to the model nested fashion (hence the name). After introducing the Laplace approximation in section 2.7.3 and the implicitly required Gaussian approximation (section 2.7.4) the density exploration strategy necessary to integrate over the full posterior approximation $\tilde{p}(\mathbf{\Theta}|\mathbf{y})$ is described in section 2.7.5.

2.7.3 The Laplace Approximation

The first step of the Laplace approximation is to apply Bayes' formula twice. In case of the hyper-parameters this leads to

$$p(\mathbf{\Theta}|\mathbf{y}) = \frac{1}{p(\mathbf{y})}\pi(\mathbf{y},\mathbf{\Theta}) = \frac{1}{p(\mathbf{y})}\frac{p(\mathbf{x},\mathbf{y},\mathbf{\Theta})}{p(\mathbf{x}|\mathbf{y},\mathbf{\Theta})} \propto \frac{p(\mathbf{x},\mathbf{y},\mathbf{\Theta})}{p(\mathbf{x}|\mathbf{y},\mathbf{\Theta})}, \tag{2.29}$$

for arbitrary $\mathbf{x} \in \mathbb{R}^n$. Consecutively, the denominator $p(\mathbf{x}|\mathbf{y},\mathbf{\Theta})$ of the right hand side of Equation (2.29) is approximated by a Gaussian $\tilde{p}_G(\mathbf{x}|\mathbf{y},\mathbf{\Theta})$. Rue and Martino (2007) show that this approximation is particularly accurate at the value $\mathbf{x}^*(\mathbf{\Theta})$ that maximizes $p(\mathbf{x}|\mathbf{y},\mathbf{\Theta})$. This gives approximation

$$p(\mathbf{\Theta}|\mathbf{y}) \approx \tilde{p}(\mathbf{\Theta}|\mathbf{y}) = \frac{p(\mathbf{x},\mathbf{y},\mathbf{\Theta})}{\tilde{p}_G(\mathbf{x}|\mathbf{y},\mathbf{\Theta})}\Big|_{\mathbf{x}=\mathbf{x}^*(\mathbf{\Theta})}. \tag{2.30}$$

In case of the

$$p(x_i|\mathbf{x}_{-i}\mathbf{\Theta},\mathbf{y}) \approx \tilde{p}_{\mathrm{LA}}(x_i|\mathbf{x}_{-i}\mathbf{\Theta},\mathbf{y}) = \frac{p(\mathbf{x},\mathbf{y},\mathbf{\Theta})}{\tilde{p}_{GG}(\mathbf{x}_{-i}|x_i,\mathbf{y},\mathbf{\Theta})}\Big|_{\mathbf{x}_{-i}=\mathbf{x}^*_{-i}(\mathbf{\Theta})}. \tag{2.31}$$

of the latent marginal posteriors. The construction of the Gaussian $\tilde{p}_G(\mathbf{x}|\mathbf{y},\mathbf{\Theta})$ is based on a sub-step of a sampling strategy brought up by Knorr-Held and Rue (2002). The same approach can be applied to the denominator $p_{GG}(x_i|\mathbf{x}_{-i},\mathbf{y},\mathbf{\Theta})$ of the latent field marginals. However, this comes with some intricacies and alternative approximations later discussed in section 2.7.6. In the following section the principle of the Gaussian approximation is conveyed.

2.7.4 The Gaussian Approximation

Assume that the observations $\mathbf{y}$ and the hyper-parameters $\boldsymbol{\Theta}$ are given, e.g. as an input to the evaluation of the density function $p(\boldsymbol{\Theta}|\mathbf{y})$ and the respective Laplace approximation. The key idea of the incorporated Gaussian approximation is to make use of the fact that the full conditional for a zero mean Gaussian $\mathbf{x}$,

$$p(\mathbf{x}|\boldsymbol{\Theta}, \mathbf{y}) \propto \exp\Big(-\frac{1}{2}\mathbf{x}^T Q\mathbf{x} + \sum_i g_i(x_i)\Big) \tag{2.32}$$

with likelihoods $g_i(x_i) = \log p(y_i|x_i)$ is often well approximated by a Gaussian distribution if the modal configuration and curvature at the mode are matched. Note that in the following we will for convenience notationally subpress all dependencies on the hyper-parameters. The approximation itself is achieved by an iterative optimization scheme alternating between optimizing for the two objectives. Assume a given mean $\boldsymbol{\mu}^{(k)}$ for the approximation, i.e. an initial guess or an estimate from the previous iteration k. One now expands the functions g_i around the respective component means μ_i^k to the second order in order to match the curvature,

$$g_i(x_i) \approx g_i(\mu_i^{(k)}) + b_i x_i - \frac{1}{2}c_i x_i^2, \tag{2.33}$$

where the $\{b_i\}$ and $\{c_i\}$ depend on the mode $\boldsymbol{\mu}^{(k)}$. The mean of the next iteration is now obtained by observing that the non quadratic terms in Equation (2.32) only depend on single components of $\mathbf{x}$. A Gaussian approximation to this hence has a precision matrix $Q + \mathrm{diag}(\mathbf{c})$ and its mean $\boldsymbol{\mu}^{(k+1)}$ is is required to solve

$$(Q + \mathrm{diag}(\mathbf{c}))\boldsymbol{\mu}^{(k+1)} = \mathbf{b}. \tag{2.34}$$

Most importantly, if $\mathbf{x}$ is a GMRF, then, due to the construction of the approximative precision matrix, this property is set forth to the approximation and respective computational properties can be taken advantage of. This optimization strategy, also known as the Newton-Raphson method or scoring algorithm and its variant the Fisher scoring algorithm (Fahrmeir and Tutz, 2001), is applied until convergence to a Gaussian with mean $\boldsymbol{\mu}^*$ and diagonal precision parameter $\mathbf{c}^*$ such that

$$p(\mathbf{x}|\boldsymbol{\Theta}, \mathbf{y}) \approx \tilde{p}_G(\mathbf{x}|\boldsymbol{\Theta}, \mathbf{y}) \propto \exp\Big(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}^*)^T(Q + \mathbf{c}^*)(\mathbf{x} - \boldsymbol{\mu}^*)\Big). \tag{2.35}$$

2.7.5 Hyper-parameter Exploration and Integration

Once the Laplace approximation of the full hyper-parameter posterior is available a computationally rational strategy for its exploration and subsequent integration is required. This is due to the fact that the previously introduced algorithmic steps are required for each function evaluation with respect to $\boldsymbol{\Theta}$. Rue and Martino (2007)

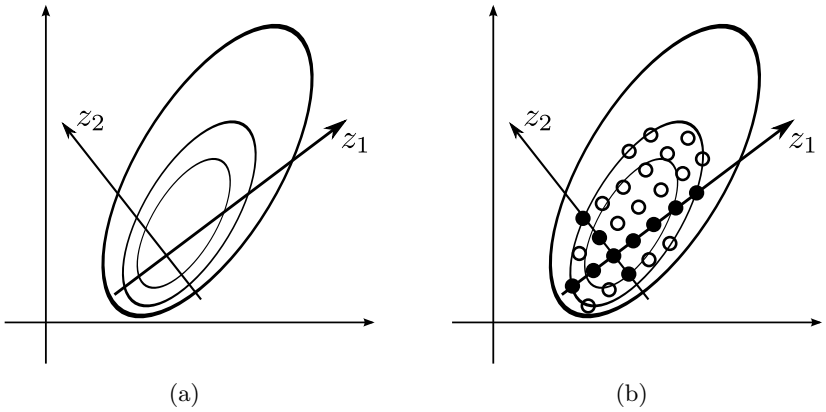


Figure 2.5: Exploration of the hyper-parameter space, adapted from Rue et al. (2009). Panel (a) shows the coordinate system of $\mathbf{z}$ with origin at the mode of the posterior and contour lines according to the respective Hessian. Each coordinate in Panel (b) is then explored independently (black markers) until a certain threshold is reached. These define the limit for the exploration of the combined coordinate directions (hollow markers).

show that the following procedure (later elucidated in more detail by Rue et al. (2009)) is accurate and particularly suitable for the exploration of $\tilde{p}_G(\boldsymbol{\Theta}|\mathbf{y})$.

Assume that $\boldsymbol{\Theta} = (\Theta_1, \ldots, \Theta_m) \in \mathbb{R}^m$, which can always be obtained by reparameterization.

1. Locate the mode $\boldsymbol{\Theta}^*$ of $\log(\tilde{p}_G(\boldsymbol{\Theta}|\mathbf{y}))$ by a quasi-Newton method using finite differences as approximation to the gradient and second derivatives.

2. Again using finite differences compute the negative Hessian matrix $H > 0$ at the mode. If the density were Gaussian, H is the respective precision and $\Sigma = H^{-1}$ the covariance matrix. The density is now explored via a parameterization $\mathbf{z}$ that standardizes $\boldsymbol{\Theta}$. That is, let $V\Lambda V^T$ be the eigen-decomposition of Σ and define
$$\boldsymbol{\Theta}(\mathbf{z}) = \boldsymbol{\Theta}^* + V\Lambda^{1/2}\mathbf{z}. \tag{2.36}$$
In case the density is Gaussian then $\mathbf{z}$ is $\mathcal{N}(\mathbf{0}, \mathbf{I})$ and hence a correction for rotation and scaling is performed.

3. For each component z_i evaluate the density at equally distant points in the positive and negative direction until a certain lower threshold, say δ_p, is met. This results in component wise numbers $s_i^{(+/-)}$ of steps into each direction. These serve as upper thresholds on the number of steps taken in the subsequent joint exploration of components of $\mathbf{z}$. A depiction of this strategy is shown in Figure 2.5.

With these exploration points at hand, two integration tasks can be fulfilled in this context.

First, the integration of $\tilde{p}_G(\boldsymbol{\Theta}|\mathbf{y})$ that is required to infer the hyper parameter marginal approximations $\tilde{p}(\theta_j|\mathbf{y})$. In this case Rue et al. (2009) recommend to compute an interpolant to $\tilde{p}_G(\boldsymbol{\Theta}|\mathbf{y})$ and then deploy a generic numerical integration scheme. The choice of the step size of the exploration scheme therefore clearly influences the accuracy of the estimated marginals. Numerical comparisons with respect to this parameter of the method can be found in Martino (2007).

The second case of integration where the full hyper-parameter posterior is required is the computation of the latent marginals $p(x_i|\mathbf{y})$ as stated in Equation (2.27). Here, no interpolant is used and the grid points of the exploration technique are used directly. The numerical integral of the marginals then reads as

$$\tilde{p}(x_i|\mathbf{y}) = \sum_k \tilde{p}(x_i|\boldsymbol{\Theta}_k, \mathbf{y}) \times \tilde{p}(\boldsymbol{\Theta}_k|\mathbf{y}) \times \Delta_k, \tag{2.37}$$

where the summation weights Δ_k are chosen to be equal and to normalize the marginals. Yet, Rue et al. (2009) propose three alternatives for the approximation of the conditional marginals $p(x_i|\mathbf{y}, \boldsymbol{\Theta})$, which will be discussed in the next section.

2.7.6 Approximations to Conditional Latent GMRF Marginals

Rue et al. (2009) propose three alternatives to the approximation of the latent GMRF marginals conditional on the data. A straight forward way, already brought up by Rue and Martino (2007), is to employ the Gaussian full posterior approximation $\tilde{p}_G(\mathbf{x}|\mathbf{y}, \boldsymbol{\Theta})$ determined during the exploration of the hyper-parameter posterior, compute the marginal distribution

$$\tilde{p}_G(x_i|\boldsymbol{\Theta}, \mathbf{y}) = \mathcal{N}(x_i|\mu_i(\boldsymbol{\Theta}), \sigma_i^2(\boldsymbol{\Theta})) \tag{2.38}$$

for each x_i and subsequently perform the integration over the hyper-parameter distribution (Equation (2.27)). The nested Laplace approximation expressed by Equation (2.31) is not used in this case. As discussed in section 2.6, this strategy is computationally very effective since $\tilde{p}_G$ is a GMRF and the marginals only depend on the respective (sparse) variable neighborhood implied by the precision matrix. Unfortunately a major drawback of this approach was shown by Rue and Martino (2007). Some of the marginals showed slight errors in location and/or a lack of skewness. Furthermore, it was shown that it is difficult to detect the x_i's for which the approximation is less accurate.

The second mean for inference on the marginals is to actually perform the nested Laplace approximation stated in Equation (2.31). This raises the problem that the Gaussian approximation $\tilde{p}_{GG}$ and the respective optimization steps (see section 2.7.4)

have to be performed for each of the latent variables. Two main ideas allow to circumvent this issue while maintaining the accuracy lost with the aforementioned marginalization approach.

- The first modification avoids the optimization step of fitting the mode of $\tilde{p}_{GG}(\mathbf{x}_{-i}|x_i, \mathbf{y}, \boldsymbol{\Theta})$ to $p(\mathbf{x}_{-i}|x_i, \mathbf{y}, \boldsymbol{\Theta})$ for each x_i. Instead, the respective modal configurations are approximated by conditioning the joint posterior $\tilde{p}_G$ on the respective x_i and taking the expectation, i.e.,

$$\mathbf{x}^*_{-i}(x_i, \boldsymbol{\Theta}) \approx \mathbb{E}_{\tilde{p}_G}(\mathbf{x}_{-i}|x_i). \tag{2.39}$$

  Again the properties of GMRFs play a significant role in this case as the conditional mean can be computed from a rank one update from the unconditional mean.

- Due to the latent GMRF structure Rue et al. (2009) argue that only those x_j should have a significant influence on the marginal of x_i that are inside a 'region of interest' $R_i(\boldsymbol{\Theta})$ closely surrounding x_i. The region itself is constructed by observing that Equation (2.39) implies

$$\frac{\mathbb{E}_{\tilde{p}_G}(x_j|x_i)}{\sigma(j)(\boldsymbol{\Theta})} = a_{ij}\frac{x_i - \mu_i(\boldsymbol{\Theta})}{\sigma_i(\boldsymbol{\Theta})} \tag{2.40}$$

  for some $a_{ij}(\boldsymbol{\Theta})$ when $i \neq j$. Membership of the region of is then determined by the threshold

$$R_i(\boldsymbol{\Theta}) = \{j : |a_{ij}(\boldsymbol{\Theta})| > 0.001\}$$

  and hence only a factorization of a sparse $|R(\boldsymbol{\Theta})| \times |R(\boldsymbol{\Theta})|$ is needed to compute $\tilde{p}_{GG}$ for a given x_i.

- As a last way to reduce the costs of the latent marginals an interpolation scheme for the density $\tilde{p}_{\text{LA}}$ is proposed. The idea is to avoid having to compute $\tilde{p}_{\text{LA}}$ in Equation (2.31) for as many different values of x_i as possible. For this purpose, the standardized marginal Gaussian approximations

$$x_i^{(s)} = \frac{x_i - \mu_i(\boldsymbol{\Theta})}{\sigma_i(\boldsymbol{\Theta})}$$

  are employed to determine the interpolation points according to the Gauss-Hermite quadrature rule. The density $\tilde{p}_{\text{LA}}$ is then represented as

$$\tilde{p}_{\text{LA}}(x_i|\boldsymbol{\Theta}, \mathbf{y}) \propto \mathcal{N}(x_i|\mu_i(\boldsymbol{\Theta}), \sigma_i^2(\boldsymbol{\Theta})) \times \exp(\text{qspl}(\text{x}_\text{i})), \tag{2.41}$$

  where $\text{qspl}(\text{x}_\text{i})$ is a cubic spline fitted to the difference to the log-density of $\tilde{p}_{\text{LA}}(x_i|\boldsymbol{\Theta}, \mathbf{y})$ and $\tilde{\tilde{p}}_G(x_i|\boldsymbol{\Theta}, \mathbf{y})$ at these points. The density is then normalized by quadrature integration.

Lastly, Rue et al. (2009) develop a third algorithm to approximate the latent marginal distributions. The so called *Simplified Laplace Approximation* $\tilde{p}_{\mathrm{SLA}}(x_i|\boldsymbol{\Theta}, \mathbf{y})$ is based on a series expansion of $\tilde{p}_{\mathrm{LA}}(x_i|\boldsymbol{\Theta}, \mathbf{y})$ around $x_i = \mu_i(\boldsymbol{\Theta})$ in order to correct the Gaussian approximation $\tilde{p}_G(x_i|\boldsymbol{\Theta}, \mathbf{y})$ for the aforementioned errors in location and skewness. Details on this approximation are given in the monographs of (Rue et al., 2009, sec. 3.2.3).

3 Dust Detection

The very basis of the work this thesis presents is the detection of dust aerosols from SEVIRI imagery. In environmental sciences a common mode of visually detecting dust from this data is to inspect linear mixtures of different SEVIRI infrared channels, the so called SFI. These mixtures project the infrared signal to red, green and blue components such that dust plumes are most prominent to the expert inspecting the image. For mathematical and computationally automated assessment of dust aerosols, however, a precise quantification of this "prominence" is required. The main goal is thus to equip each region represented by a pixel in the imagery with a scalar quantity characterizing evidence for dust presence.

This chapter describes the development of a probabilistic framework expressing evidence for dust presence via a flexible Bayesian hierarchical model (BHM) (see Section 2.4 for details on BHMs). Inference in this framework is straight forward using the integrated nested Laplace approximations approach described in Section 2.7. Section 3.1 gives an overview about the specifications of SEVIRI data, the mathematical definition of the SFI and finally describes a thresholding scheme for dust detection developed by Ashpole and Washington (2012). Together with a baseline GLM (see Section 2.5) called linear discriminant analysis (LDA) this scheme serves to evaluate the proposed probabilistic framework in terms of detection accuracy.

Section 3.2 presents first steps towards probabilistic dust detection, which are taken from Bachl and Garbe (2012). Here, LDA is embedded into BHM and subsequently extended to projection coefficients modeled as functions over background estimates. The latter represent SEVIRI imagery under pristine sky conditions. As these functions are unknown and have to be inferred from data, the method is named latent projection functions (LPFs) from here on.

One drawback of linearly projecting even mildly noisy data like SEVIRI imagery is that no non-linearity can be employed to dampen signal disturbances. The noise carried over to the projected quantity indicating dust can then hamper further processing like estimation of the transport process. Section 3.3, adapted from Bachl et al. (2012), displays our work to overcome this shortcoming of LPFs. A BHM is examined that integrates the SEVIRI data as a part of the domain of the (non-linear) latent functions. This method, named latent signal mappings (LSMs), yields less noisy spatio-temporal dust indication and, as will be shown in Chapter 4), aids inference on the underlying atmospheric transport process.

The last section of this chapter (3.4) is dedicated to improvements in terms of LSM's stability over day time and dependency on background estimates. By replacing the latter with surface emissivity estimates a major drawback of this approach is circumvented (Bachl et al., 2013a). The formerly proposed method of background radiation retrieval is not capable to yield stable results for times with low solar altitude, e.g. at early mornings where dust storms are known to emerge frequently. The section is concluded by an examination of a cross validation study taken from Bachl et al. (2013b). The study shows that LSM with emissivity estimates yields a detection accuracy that is, in particular with respect to day time, significantly favorable over approaches like thresholding and LDA.

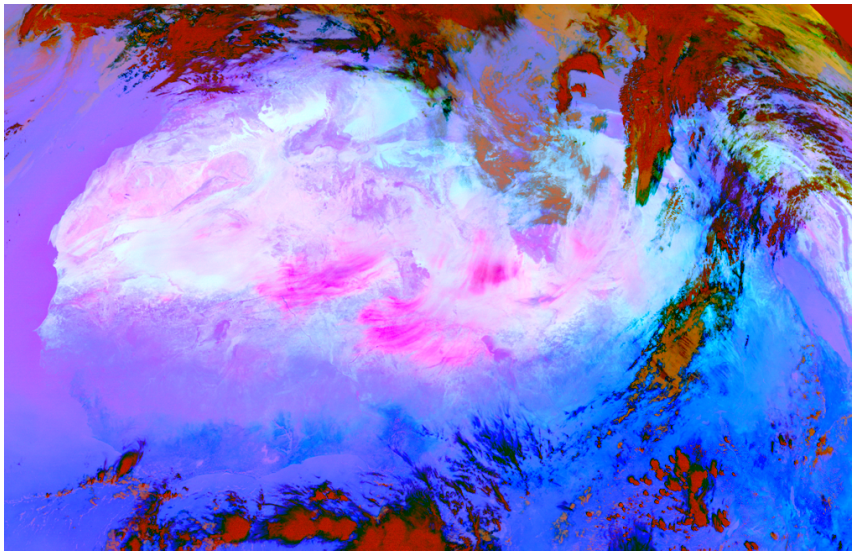
3.1 Preliminaries

3.1.1 SEVIRI Data and Falsecolor Representation

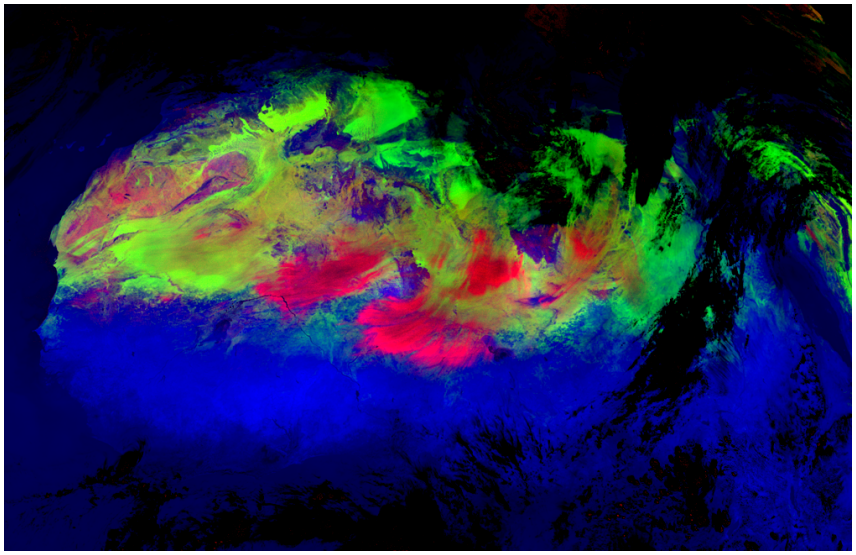
The SEVIRI instrument resides aboard the Meteosat-9 satellite launched on December 21, 2005 in a joint effort of the European Organization for the Exploitation of Meteorological Satellites (EUMETSAT) and the European Space Agency (ESA). Being an integral part of the payload of the Meteosat Second Generation (MSG) series of platforms dedicated to environmental data collection, SEVIRI measures electromagnetic radiation at 12 different spectral windows spanning from visible to infrared frequencies (Schmetz et al., 2002). With Meteosat-9 residing at 0 degrees of latitude, 0 degrees of longitude and a height of approximately 36 km it provides measurements for up to 80 degrees of deviation from nadir where it has a resolution of about 3×3 km. In combination with the per-image scan time of 12 minutes and three minutes of calibration this results in a 3712×3712 pixel imagery every 15 minutes.

With respect to radiative remote sensing, the most dominant effect of dust aerosols is to filter the infrared radiation leaving the terrestrial surface in a frequency dependent fashion. This phenomenon is reflected by the 12.0 μm, 10.8 μm and 8.7 μm brightness temperature (BT) measurements of the SEVIRI instrument, which we will abbreviate by $BT_{12.0}$, $BT_{10.8}$ and $BT_{8.7}$, respectively. For example, it is well known that in the presence of dust aerosols the difference $\Delta T_{BR} = BT_{12.0} - BT_{10.8}$ increases while $\Delta T_{BG} = BT_{10.8} - BT_{8.7}$ decreases (Schepanski et al., 2007). This connection, also known as split window technique, results in popular operative products such as the SFI (see Figure 3.1(a)) which defines the red (R), green (G) and blue (B) channels of the visualization as

$$\begin{aligned} R &= (\Delta T_{BR} + d_R)/s_r, \\ G &= ((\Delta T_{BG} + d_G)/s_G)^\gamma, \\ B &= (BT_{10.8} - d_B)/s_B, \end{aligned}$$



(a)



(b)

Figure 3.1: SEVIRI falsecolor schemes for dust detection. Panel (a) shows dust event on January 17, 2010 at 12h GMT over northern Africa via the non-linear SFI scheme. Panel (b) shows the same event via the linear version $SFI_{\gamma=1}$ described in section 3.1.1. The variants show dust as (a) purple and (b) red regions. Water clouds are (a) black to dark red and (b) black. In (a) the surface of the earth is represented by blue to cyan and in north western Africa with slightly purple colors. Differences in the vegetation are easier to derive from (b), where blue areas are more vegetated than green area.

where the linear rescaling parameters

$$\begin{aligned} d_R &= 4K, & s_R &= 2K + 4K, \\ d_G &= 0K, & s_G &= 6K, \\ d_B &= 261K, & s_B &= 289K - 261K, \end{aligned}$$

map the data to the interval $[0, 1]$ such that changes due to dust activity are most noticeable during on-screen inspection by experts. The parameter γ is set to 0.4 and K denotes the unit of brightness temperature in Kelvin. A detailed discussion on this and related falsecolor products was published by Lensky and Rosenfeld (2008). Purely linear schemes with $\gamma = 1$ are common as well (see, e.g., Schepanski (2009)) and can be visually advantageous depending on the medium the data is displayed on. For instance, using the parameters

$$\begin{aligned} d_R &= -0.08K, & s_R &= 0.35K, \\ d_G &= -0.3K, & s_G &= 0.08K, \\ d_B &= -5K, & s_B &= 0.02K, \end{aligned}$$

more contrast is given to surface features like vegetation. Figure 3.1(b) depicts the resulting imagery as opposed to the default SFI shown in Figure 3.1(a). Throughout this thesis, whenever this falescolor parameterization is employed, it will be denoted by $\text{SFI}_{\gamma=1}$. If data $\{D_{i\in\{1,2,3\}}\}$ other than the brightness temperature channels is processed according to a SFI scheme this will be denoted by $\text{SFI}(D_1, D_2, D_3)$ and $\text{SFI}_{\gamma=1}(D_1, D_2, D_3)$, respectively.

3.1.2 Split Window and Tresholding Techniques

Recently, Ashpole and Washington (2012) proposed an extended thresholding scheme for dust detection based on the split window technique. They combine four different thresholds given by

$$\begin{aligned} \Delta T_{BR} &> 0K, & (3.1) \\ \Delta T_{BG} &< 10K, & (3.2) \\ BT_{10.8} &< 285K, & (3.3) \\ \Delta T_{BR} - M &< -2K, & (3.4) \end{aligned}$$

where M is a two-week cloud masked rolling mean of ΔT_{BR}. Alongside requiring the fixed conditions given in Equations (3.1) and (3.2) in order to flag a pixel to contain dust they introduce two additional requirements. Since the blue channel is generally saturated in the presence of dust while the occurrence of clouds lowers its brightness, the threshold $BT_{10.8} < 285K$ in Equation (3.3) removes artifacts coming from the latter. The last threshold is data dependent and serves two purposes. By requiring Equation (3.4) to hold it rules out false positive dust detections where clouds are

present and over regions where the red channel is close to saturation even under pristine conditions.

Ashpole and Washington (2012) perform an extensive study comparing their SEVIRI based approach to ground measurements of AOD at different photometer sites of the Aerosol Robotic Network (AERONET) and the absorbing aerosol index (AAI) derived from the ozone monitoring instrument on board the Aura satellite[1]. They conclude a general accordance with these means of dust detection but also point out that their method suffers from a decreased detection accuracy in the early mornings and late evening due to a "decrease in the spectral contrast between cooler surfaces and the dust layer".

Day time is not the only critical factor for dust detection from SEVIRI data. Related work (Brindley et al., 2012) employs an extensive simulation of SFI data given different atmospheric conditions, seasons of the year, day time, vegetation cover as well as other influencing factors like satellite viewing angle. Figure 3.2 summarizes two of their main results, which play a mayor role in the context of this thesis.

- The vegetation cover of the earth determines which of the SFI channels is most expressive in terms of dust detection. Under pristine conditions the green channel is comparably saturated for barren surfaces while the red channel is desaturated. Covered by dust the green channel is desaturated significantly and only a minor saturation of the red channel occurs. With increasing vegetation cover this effect is reversed. Vegetated surfaces show a low and high saturation of the green and red channel, respectively. Dust then leads to a strong reduction of the red channel intensity while the green channel is less affected.

- The behavior with respect to day time is slightly more complex and also depends on the season of the year. With increasing day time a general shift of the appearance under pristine conditions occurs towards less red and more green saturation. This shift is even stronger in case of dust presence such that the two conditions are less visually discriminable. According to the simulations study this behavior is more pronounced at summer time. During winter time slight ambiguities are visible, especially for regions with a medium green saturation of around 5 K.

It should also be noted that the study analyzes the SFI appearance in dependence of the dust optical depth τ_{10} at $10\mu m$ in steps of 0.1 from 0.0 to 1.0. The corresponding transition in the SFI color space is very smooth in the red an green components (see the moderate curvature of the arrows in Figure 3.2). A sufficiently smooth function that depends on the vegetation level and/or day time and maps the SFI to a scalar, e.g. a probability of dust occurrence, can thus be interpreted as an approximation to

[1] AERONET and Aura are maintained by the National Aeronautics and Space Administration (NASA)

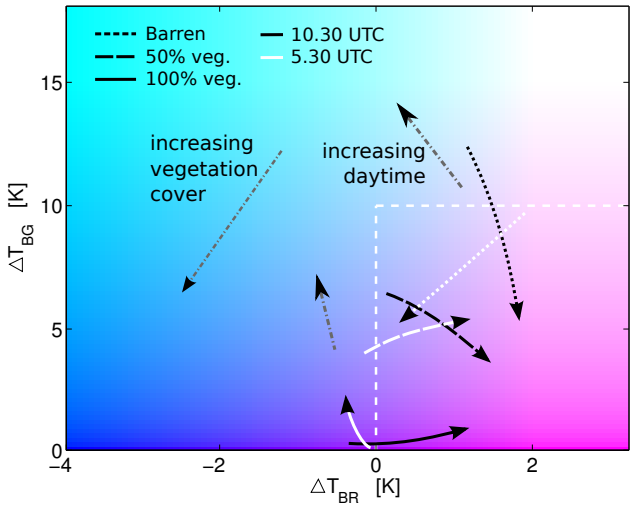


Figure 3.2: Simulated dust appearance in SEVIRI falsecolor imagery (December atmospheric conditions), adapted from Brindley et al. (2012). Assuming a fully saturated blue channel the coloration of the axis system shows the SFI appearance depending on the red (abscissa) and green (ordinate) channel saturation. The general tendency of pristine pixels to change appearance with varying vegetation cover and day time is depicted by the grey arrows. Black and white arrows depict the SFI appearance transition from an assumed dust optical depth τ_{10} of 0 to 1 at a wavelength of $10\mu m$ and in dependence of the surface vegetation and day time. The dashed white line shows the first two dust detection thresholds (Equations 3.1 and 3.1) proposed by Ashpole and Washington (2012).

the optical depth τ_{10}. For a preview of how actual (hand labeled) data is distributed in the SFI space see also Figure 1.3 in the introduction of this thesis.

3.1.3 Linear Discriminant Analysis

Linear discriminant analysis is a simple but nonetheless powerful classification technique. As the name implies, LDA imposes a linear decision surface on the space of covariates (see figure 3.3). The general idea of LDA is to find projection coefficients $\mathbf{w}$ of the covariates that maximizes the generalized Rayleigh quotient

$$R(\mathbf{w}) = \frac{\mathbf{w}^T S_B \mathbf{w}}{\mathbf{w}^T S_W \mathbf{w}}. \tag{3.5}$$

In this equation, S_B is outer product of the difference of the two class means $\mathbf{m}_1$ and $\mathbf{m}_2$ and S_W is the sum of the scatter matrices

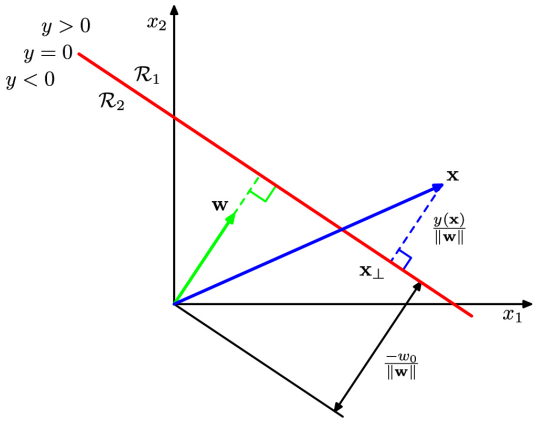


Figure 3.3: Geometry of a linear decision boundary defined by LDA, adapted from Bishop (2006). The depicted LDA decision boundary (red) separates the two-dimensional space spanned by x_1 and x_2 into two regions $\mathcal{R}_1$ and $\mathcal{R}_2$. The vector $\mathbf{x}$ (blue) is classified according to the sign of its distance $y(\mathbf{x})/||w||$ from the decision boundary being perpendicular to the projection vector $\mathbf{w}$.

$$S_i = \sum_{\mathbf{x} \in \mathcal{D}_i} (\mathbf{x} - \mathbf{m}_i)^T (\mathbf{x} - \mathbf{m}_i), \tag{3.6}$$

where $i \in \{1, 2\}$ is the class index and $\mathbf{x} \in \mathcal{D}_i$ are the samples of the two classes. If $\mathbf{w}$ fulfills this condition, the one-dimensional projection $\mathbf{w}^T\mathbf{x}$ has the property that it maximizes the difference of the projected means while minimizing the corresponding within-class variances. It can be determined by $\mathbf{w} = (S_1 + S_2)^{-1}(\mathbf{m}_1 - \mathbf{m}_2)$. Finally, if the constant w_0 is set to $-\mathbf{w}^T(\mathbf{m}_1 + \mathbf{m}_2)/2$, the hyperplane defined via the equation

$$y(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + w_0 = 0 \tag{3.7}$$

can be shown to be the optimal decision boundary if the two sample sets are distributed according to normal distributions with equal but unknown covariance matrices. Nonetheless, LDA also has its disadvantages. While the linear discrimination of low dimensional covariate spaces is robust against overfitting, the underlying statistics of LDA can be severely distorted by outliers. Furthermore, in particular the within-class scatter matrices suffer from the curse of dimensionality. If the number of samples is low compared to the dimensionality of the covariate space, these matrices often overestimate the distortion caused by the actual covariance matrices.

3.2 Latent Projection Functions

When visually examining SFI imagery two main features allow to recognize dust activity: motion and coloration. This section, which was published by Bachl and Garbe (2012), presents first steps towards the application of a BHM (see Section 2.4 for details on BHMs) to the problem of dust detection from features in the SFI space. The approach focuses on solving one of the most apparent difficulties that comes with detection solely based on the color space. Although it is known that dust presence increases the red and decreases the green channel intensity, the extent of this effect depends strongly on the radiative properties of the respective earth surface being covered by a pixel of the imagery. A projection of the data to a one-dimensional space that represents an indicator for dust activity should therefore depend on such surface properties. The following sections illustrate how this requirement is realized via two main building blocks. In section 3.2.2 a technique is explained that allows for an approximation to the SFI under clear sky conditions. Thereby the radiative properties of the surface represented by each pixel under the assumption that neither dust nor water clouds contaminate the signal are estimated. These features are then utilized to define a model for the detection itself.

As explained in Section 3.2.3, the core idea is to set LDA forth to a Bayesian model wherein the projection coefficients are functions over the space of terrestrial surface properties. Section 3.2.4 then goes into experimental details about two qualitative case studies that serve to evaluate the method. Finally, respective conclusions are drawn in Section 3.2.5.

3.2.1 Notation

Let $\Omega \subset \mathbb{R}^2$ denote the image domain and assume the presence of a series of images obtained over the time interval $[0, T]$. The goal is to determine a probabilistic dust indicator d_{xyt} with location $(x, y) \in \Omega$ at time $t \in [0, T]$ such that $d_{xyt} = 1$ if there is evidence of a dust plume covering the location and $d_{xyt} = 0$ otherwise. This assessment is made on the basis of the observation vector $\boldsymbol{I}_{xyt} = (I_{1xyt}, I_{2xyt}, I_{3xyt})$ where the three components of $\boldsymbol{I}_{xyt}$ correspond to the red, blue and green channels of $\text{SFI}_{\gamma=1}$.

3.2.2 Estimation of Background Appearance

A crucial aspect of dust detection via LPF is that an estimate of the background radiation is taken into account, i.e. an approximation to the earth's brightness temperature under clear sky conditions. The basic idea to obtain this estimate works as follows. Dust plumes as well as water vapor (clouds) decrease the brightness temperature measured by SEVIRI at that particular location. On the other hand,

the most significant contributor to changes in the measured infrared radiation is the thermal energy provided by the sun and varying with its angle and hence day time. This influence can approximately be excluded by considering a sequence of measurements over multiple days but at a particular time during the day only. Of these measurements the most intense one is then very likely to come from a day where that location was not covered by dust or a cloud. Given a SEVIRI brightness temperature sequence $BT_{\lambda,x,y,t}$ for time $t \in [0, T]$, spatial location $(x, s) \in \Omega$ and frequencies $\lambda \in \{12, 10.8, 8\}$ this leads to the maximum-intensity criterion

$$A_{\lambda,x,y,t} = \max_{\hat{t}} \sum_{\lambda} BT^2_{\lambda,x,y,\hat{t}} \quad \text{where} \quad \hat{t} \in \{t + 96k | k \in \mathcal{N}\}. \tag{3.8}$$

A background estimate $\mathbf{A}_{xyt} = (A_{12,x,y,t}, A_{10.8,x,y,t}, A_{8.7,x,y,t})$ for fixed t is thus a a combination of data from different days. An example for the outcome of this procedure is given in Figure 3.4, which shows a large scale dust event over northern Africa together with the estimated background for the same area. Both are depicted using the $\text{SFI}_{\gamma=1}$ scheme. The same figure also gives an example for a region in Africa for which the background estimate is essential in terms of dust detection as the radiative properties are similar to that of dust events. Note that the background estimation procedure is, however, independent of the detection techniques explained in the following paragraphs. It is therefore possible to employ other techniques to determine this estimate, e.g. applying the method introduced by Liu et al. (2012) to an extended data set.

3.2.3 Latent Projection Functions

The LPF method models the probability $\mathbb{P}(d_{xyt} = 1)$ of a pixel containing dust activity via a binomial distribution as the topmost level of a BHM. Its mean is defined via the logisitic sigmoid inverse link function (see Section 2.5 for details on links functions and GLMs) of a linear predictor $\eta(x, y, t)$:

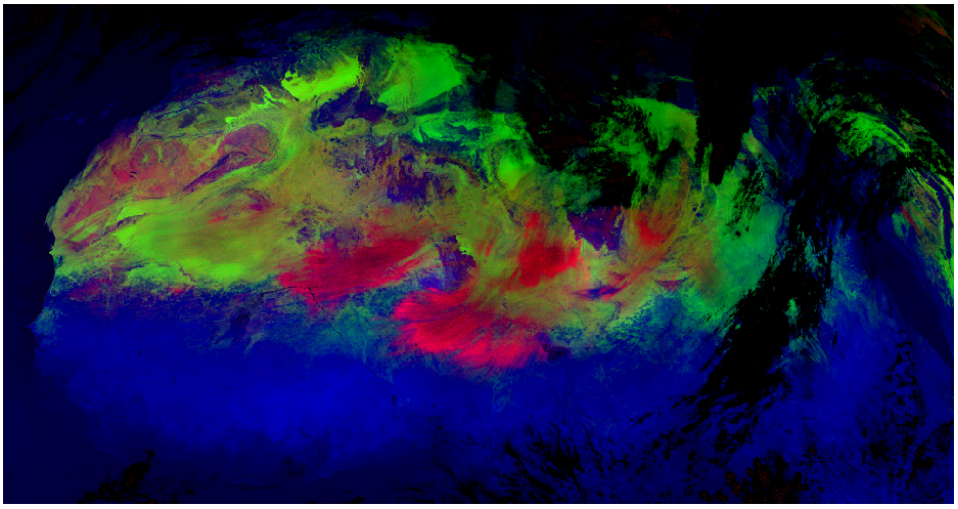
$$\mathbb{P}(d_{xyt} = 1) = 1/(1 + \exp[-\eta(x, y, t)]). \tag{3.9}$$

This predictor is an extension of what is usually employed in simple logistic regression schemes like LDA, where projection coefficients r_i and an offset q are determined such that the sign of

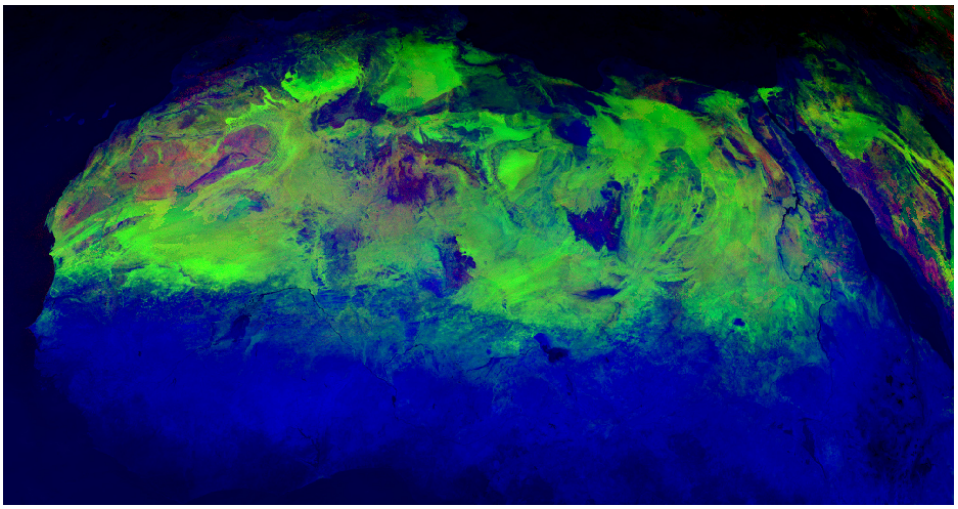
$$\hat{\eta}(x, y, t) = q + \sum_{i=1}^{3} I_{ixyt}\, r_i \tag{3.10}$$

serves as a label for the dust content of a particular location. In LPFs, however, the projection coefficients and intercepts are functions of the background estimate $\boldsymbol{A}_{xyt}$:

$$\eta(x, y, t) = \sum_{i=1}^{3} I_{ixyt}\, f_i^1(A_{ixyt}) + f_i^2(A_{ixyt}). \tag{3.11}$$



(a)



(b)

Figure 3.4: Background estimation via the maximum-intensity criterion. Panel (a) shows a dust event over central northern Africa on January 17, 2010 at 12h GMT. In Panel (b) the background appearance according to the maximum intensity criterion is shown. Note that there is a region in north west Africa (approximately Morocco) that has an appearance that is very similar to that of dust but is not contaminated.

The second and third levels of the BHM are prior distributions on the latent functions f_i^1 and f_i^2 and their parameters, respectively. The functions f_i^1 and f_i^2 are modeled semi-parametrically by binning each component of $\boldsymbol{A}_{xyt}$ into 100 distinct bins taken over the range of each component over the image domain Ω. These functionals are then modeled as second-order random walks (RW2s). That is

$$\begin{aligned} f_i^1 &\sim \mathcal{N}_{100}(0, Q_{\mathrm{RW2}}(\boldsymbol{\Theta}_i)), \\ f_i^2 &\sim \mathcal{N}_{100}(0, Q_{\mathrm{RW2}}(\boldsymbol{\Xi}_i)), \end{aligned}$$

where Q_{RW2} is set up such that the second-order forward differences are independent normals

$$\begin{aligned} \Delta^2 f_{i,j}^1 &\sim \mathcal{N}(0, 1/\kappa_i) \\ \Delta^2 f_{i,j}^2 &\sim \mathcal{N}(0, 1/\iota_i) \end{aligned}$$

and the parameters $\boldsymbol{\Theta}_i$ and $\boldsymbol{\Xi}_i$ are given independent log Gamma priors for the increment precisions κ_i and ι_i, respectively (see Section 2.5 for details on the Gamma distribution). A detailed discussion of this model is given in Section 3.4 of the monographs by Rue and Held (2005). Note that in Bachl and Garbe (2012) these models where stated as *continuous random walks* with 250 bins. Due to an error in the implementation, however, a fallback RW2 model with 100 bins was used instead.

3.2.4 Experiments & Results

This section presents two out of three experiments described by Bachl and Garbe (2012). These two concern the potential of the LPF to detect dust from SEVIRI data and the correlation of the spatial dust indicator with AOD measurements. The third experiment represents a first step to employ the dust indicator to infer the underlying transport process (Chapter 4). Details can be found in Section 4.2.

In order to induce posterior dust indication probabilities for a given SEVIRI imagery inference on the posteriors of the the latent functions f_i^1 and f_i^2 as well as their parameters is required. For that purpose the likelihood is steered by binary observation labels set in the course of a visual inspection of the data shown in Figure 3.5. Motivated by the exploratory character of this study all of the imagery is interpolated to one third of the original resolution to reduce the computational effort. Posterior inference as well as hyper parameter integration are performed using the INLA technique (see Section 2.7 for details on INLA)).

The first experiment yields a qualitative judgment of the expressive power of the linear predictor η in terms of indicating presence of dust. Given the above-mentioned labeled data and respective model posteriors, Figures 3.6(a) and 3.6(b) show a comparison between a dust event in $\mathrm{SFI}_{\gamma=1}$ representation and the positive part of η.

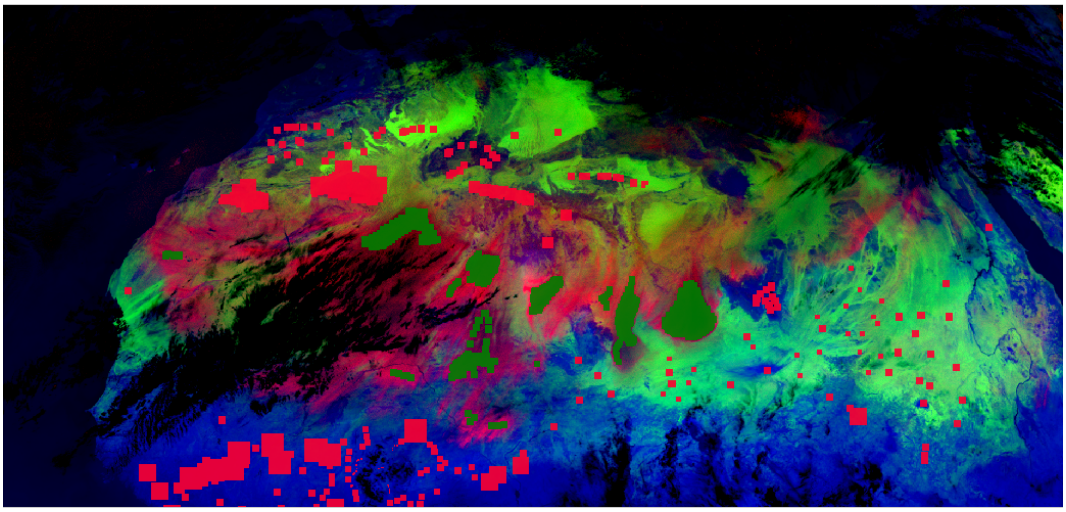


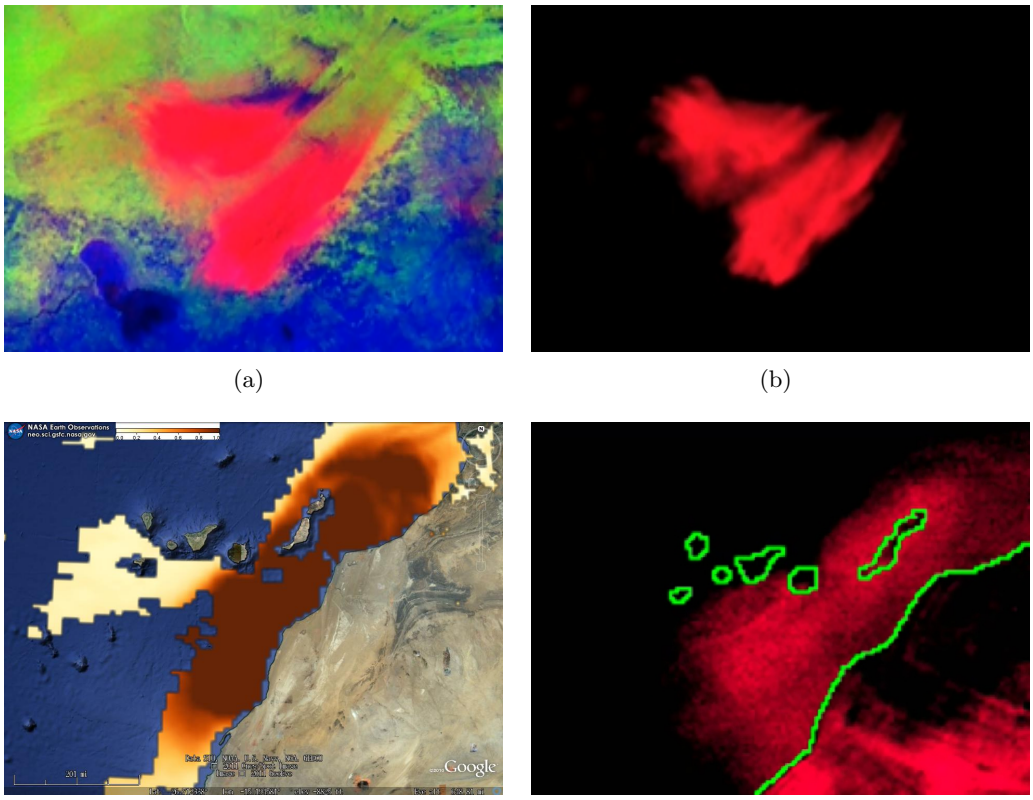
Figure 3.5: Data labels used for dust detection by LPFs. Dust and pristine regions are marked by green and red overlays, respectively.

Note that this positiveness implies that according to the model the probability of the given pixel being dusty is above 0.5. A respective decision rule would thus judge that in this case $d_{xyt} = 1$. The most important message these figures convey is that the linear predictor clearly summarizes what a visual inspection of the SFI scenery suggest. At the border of the imagery no dust is visible. Accordingly, the spatial plot of η suggests the same by taking low values. In the center of the SFI a dust plume with a faint middle line is clearly visible. The same behavior can be seen from the linear predictor. Both also show shows faint dust at the boundary of the plume, which suggests that also the optical depth of the plume is at least approximately represented by the linear predictor.

Another objective of our work is to determine areas to which the dust is transported to. This requires that the approach is capable to deal with dust passing from continental areas to those above sea. As the background variation in the latter case is rather small we reduce the latent functions to single point estimates, which results in a Bayesian logistic regression. Figure 3.6(d) shows the positive part of the corresponding linear predictors above sea and land for a dust event with a plume that passes the Canaries. For means of comparison Figure 3.6(c) displays the average AOD on that day derived from measurements of the MODIS instrument aboard the Terra platform[2]. Although the AOD data is only available for particular regions and above water a clear correspondence with the values of the linear predictor can be seen. Unfortunately direct access to the AOD data was not available at the time these experiments were conducted, preventing a quantitative comparison. It should

[2]Public access to this data is available via the NASA earth observations website http://neo.sci.gsfc.nasa.gov/

also be noted that, as opposed to AOD, the SFI only cumulates 12 minutes of dust presence, which explains the slight differences in the AOD and linear predictor.



(a) (b)

(c) (d)

Figure 3.6: Comparison of LPF linear predictors to SFI data and MODIS AOD measurements. In (a) and (b) a dust plume and the corresponding linear predictor are shown. The dust event occurred on January 18, 2010 close to the Bodélé depression in Chad. Panel (c) displays the daily average AOD of a dust plume around the Canaries. The linear predictor for this event as well as the coast line of western Africa and the Canaries is depicted in (d).

3.2.5 Conclusion

Our results show that the proposed classification method and in particular the incorporated linear predictor is a promising approach to monitor dust activity. It appears to be powerful in terms of recovering the structure of a dust plume from the multispectral signal of SEVIRI and suggests qualitative similarity to AOD measurements.

It should be mentioned that at the current stage the linear predictor of our method does not necessarily reflect a physical quantity like the mass of dust or AOD in a vertical column. Corresponding relations are still to be evaluated. However, this purely probabilistic approach seems to be a fruitful counterpart to existing physical or empirical models for which it might provide parametric input or could be validated against.

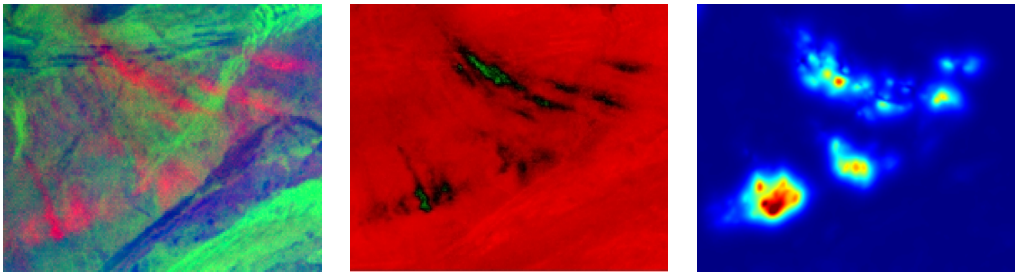
3.3 Latent Signal Mapping

In the previous sections it has been shown how a BHM (Section 2.4) can be employed to generate a linear predictor from SEVIRI data that serves as a spatio-temporal indicator for the presence of dust. The associated publication of Bachl and Garbe (2012) also describes a first approach to determine the trajectory of a dust plume by means of computing the *optical flow* from the thresholded dust predictor. Transporting back the predictor according to this flow field can then serve to point out spatial regions that are likely to be the source of the plume (see Section 4.2 for details on this topic). The thresholding serves the purpose of reducing the contribution of the predictor to the estimated flow in regions where no dust is present. This is necessary since the predictor is a background dependent linear mixture of infrared channels and thus carries over the noise incorporated by the signal.

This results in two interwoven problems. On the one hand, as shown in Figure 3.7(a) and 3.7(b), faint dust is easily underestimated by the predictor. In order to capture the motion of the dust, e.g. to provide further evidence via the presence of motion (see Figure 3.7(c) for an example) or to carry out the source estimation procedure, a lower threshold is then needed. On the other hand, lowering the detection threshold is likely to induce false positives which then interfere with the source estimation procedure.

In this contribution we extend our previous approach in the sense that the dust predictor is not a linear projection but a general mapping of the signal. The method is introduced in Section 3.3.1 and referred to as latent signal mapping (LSM).

In Section 3.3.2 we show that, in comparison with linear approaches, LSM increases the temporal signal-to-noise ratio of the predictor. That is, it decreases the variance of the respective forward differences in durst free regions while dust contaminated regions still exhibit distinguishable activity. Additionally, our method also yields a higher expressiveness of the corresponding optical flow compared to estimation based on multiple channels or linear projections thereof. This result is postponed to be presented in Section 4.3. A brief conclusion regarding LSM is then given in Section 3.3.3.



(a) (b) (c)

Figure 3.7: Detection of a faint dust plume. Panel (a) shows a dust event on May 4, 2007 which leads to low (black) or even negative values (red) of the dust predictor shown in (b). The corresponding normalized magnitude of the optical flow is depicted in Panel (c) and can serve as additional evidence for dust presence.

3.3.1 Methods

The first BHM stage of latent signal mapping (LSM) models dust presence ($d_{xyt} = 1$) and absence ($d_{xyt} = 1$) using a Binomial distribution parameterized by the linear predictor $\eta(x, y, t)$:

$$\mathbb{P}(d_{xyt} = 1) = 1/(1 + \exp[-\eta(x, y, t)]).$$

For details on the Binomial distribution and the respective GLMs see Section 2.5. Using LPFs the predictor is given by

$$\eta(x, y, t) = \sum_{i=1}^{3} I_{ixyt}\, f_i^1(A_{ixyt}) + f_i^2(A_{ixyt}), \tag{3.12}$$

where I is the SEVIRI SFI imagery, A is the respective estimate of the background radiation and the f_i^1 and f_i^2 are semi-parametrically modeled functions. This projection carries over noise in the SFI signal in a linear fashion, which can hamper consecutive processing techniques like optical flow estimation. As a remedy we propose to shift the SFI to be a part of the domain of the latent functions such that

$$\eta(x, y, t) = \sum_{i=1}^{3} h_i(A_{ixyt}, A_{ixyt} - I_{ixyt}). \tag{3.13}$$

As with LPFs the functions h_i are modeled semi-parametrically by binning each component of $\mathbf{A}_{xyt}$ and $\mathbf{A}_{xyt} - \mathbf{I}_{xyt}$ into 100 distinct bins taken over the range of each component and given the same training set used in Section 3.2.4. Each h_i is modeled as a two dimensional conditional auto-regression (CAR) GMRF (see Rue

and Held (2005) for details) with

$$p(h_i(j,k)) \propto \exp\Big(-\rho_i \sum_{(l,m)\sim(j,k)} \big(h_i(l,m) - h_i(j,k)\big)^2 \Big), \tag{3.14}$$

where "$\sim$" denotes the four nearest neighbors on the two dimensional discretization grid of $A_{ixyt} \times (A_{ixyt} - I_{ixyt})$. We assign independent log Gamma priors to the unknown precisions $1/\rho_i$ that steer the local squared deviations and thus the smoothness of the modeled function (see Section 2.5 for details on the Gamma distribution). Posterior inference was performed using the INLA technique described in Section 2.7.

3.3.2 Experiments & Results

As we focus on faint dust with a weak and/or noisy signal we inspect temporal sequences of two events that are characteristic in that sense. Event A, shown in Figure 3.8(a), is a five hour sequence of multiple faint plumes with a very smooth trajectory. Event B, shown in Figure 3.8(b), is composed of two hours of multiple small plumes coming from comparably close but distinguishable source areas. Figure 3.8 displays the normalized standard deviations of the temporal forward differences for the $\mathrm{SFI}_{\gamma=1}$ representation and the linear predictors of LDA and LSM. We omit results for the LPF method since they are very similar to those of LDA. For both events it is obvious that the activity of the predictor of LSM is considerably more expressive than for $\mathrm{SFI}_{\gamma=1}$ and LDA when comparing dusty and dust free regions. Also, in case of event B, for which even visual discrimination between dusty and non-dusty regions is challenging, the forward differences seem to be an appealing feature for the detection of the relatively small dust sources.

3.3.3 Conclusion

The results show that the proposed method of LSM is clearly to be preferred over the LPF approach to detecting dust aerosols in multispectral data. In particular the propagation of the signal intrinsic noise to the scalar indicating the presence of dust is strongly reduced. We are also able to point out that the LSM temporal forward differences are comparably expressive in terms of indicating early dust activity. This is a potentially supporting mode in terms of tracing dust back to its source area, which is a particularly desirable goal from the perspective of environmental sciences.

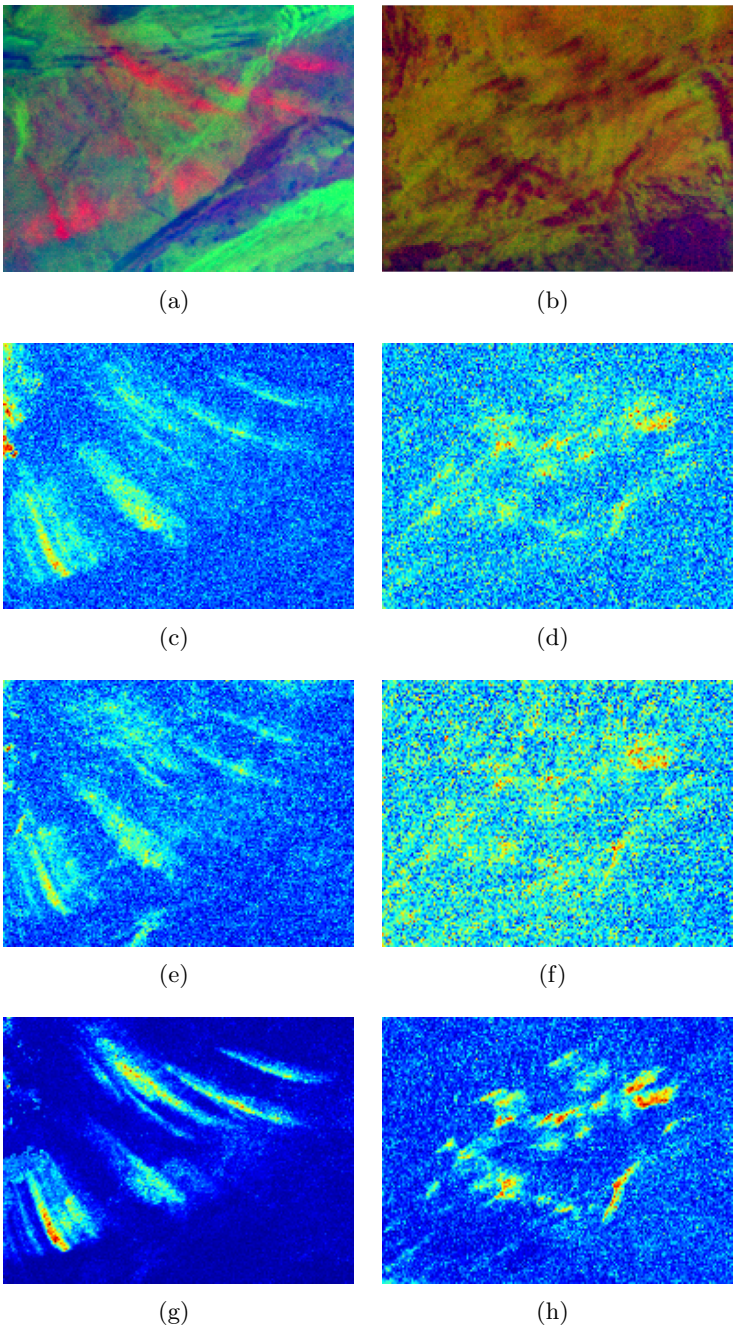


Figure 3.8: A comparison of spatially normalized standard deviations calculated from the temporal forward differences of dust events on May 4, 2007 (Panel (a)) and January 17, 2010 (Panel (b)). Panels (c) and (d) refer to the mean deviation of the falsecolor representation $\mathrm{SFI}_{\gamma=1}$, (e) and (f) to the predictor of LDA and (g) and (h) to the LSM method.

3.4 Latent Signal Mapping with Surface Emissivity

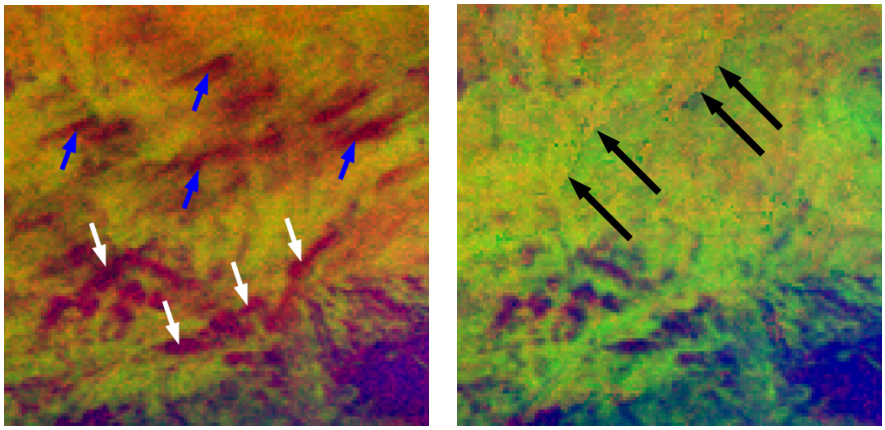
The Earth's brightness temperature under clear sky conditions varies significantly depending on the surface characteristics of sea versus land, among different vegetation zones as well as along time during a day. As Figure 3.9(a) shows, this leads to difficulties in the differentiation between some pristine areas and those contaminated with dust. Especially during the well known peak time of dust emission (8-10 a.m.) the intense change of infrared radiation induced by the rising sun complicates the detection of dust aerosols.

In order to facilitate this differentiation, our previous approaches used a maximum-intensity criterion (see Section 3.2.2) to estimate the appearance of the background, i.e., the earth's radiation as measured by SEVIRI under clear sky condition. While this technique provides impressive results in filtering out dust and cloud contaminated areas it has two undesirable side-effects. It spatially combines SEVIRI signals from different days. Inter-daily changes other than clouds and dust, e.g. air temperature and cumulative thermal energy effects, thus lead to even visually perceivable differences in the radiation received from the different combined regions. Due to a second shortcoming of the method these edge effects are particularly strong during the early morning and late afternoon (approximately before 8:30 a.m. and after 4:30 p.m.). An example of such an edge artifact is shown in in Figure 3.9(b). Here, the intrinsic assumption of radiation decrease induced by aerosols is not necessarily met due to low surface temperatures. Dust contaminated pixels can then enter the background estimate and lead to an underestimation of dust presence.

In the following sections a remedy to these issues is elucidated, based on the respective publications of Bachl et al. (2013a) and Bachl et al. (2013b). By replacing the anomaly detection term $\mathbf{A} - \mathbf{I}$ of differences between the background estimate $\mathbf{A}$ and the SFI imagery $\mathbf{I}$ (see Equation (3.13)) with monthly average emissivity estimates the LSM method evolves into an extremely powerful tool for dust detection. The respective methodology as well as the employed emissivity estimates are presented in Section 3.4.1. Thereafter, an extensive cross validation study comparing the method with LDA and thresholding techniques is presented in Section 3.4.2 and concluded upon in Section 3.4.3.

3.4.1 Methods

Each physical object with a temperature above absolute zero emits thermal radiation. The emissivity of this object is a relative quantity comparing the object with the hypothetical construct of a *black body*. A black body defines the maximum of thermal energy an object is able to emit at a particular temperature and frequency. Due to Kirchhoff's law, which states that emissivity and absorptivity are proportional, the emissivity of an object is also directly connected to the visual and infrared appearance



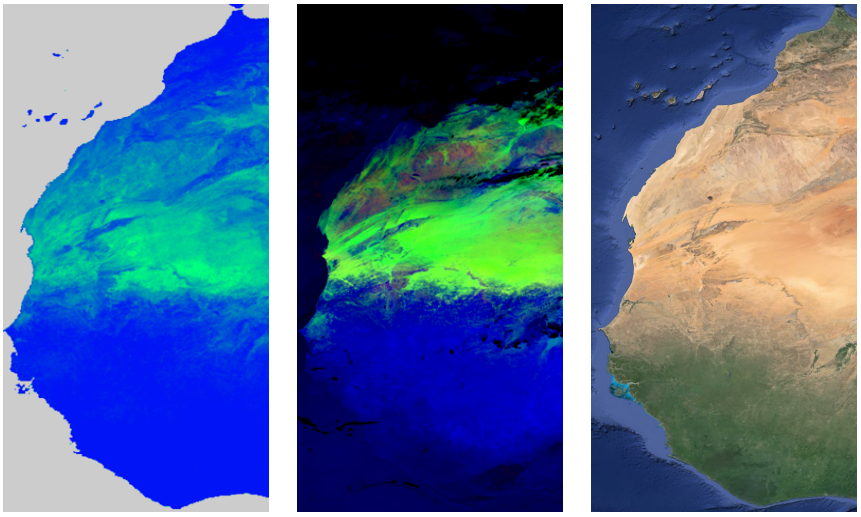
(a) (b)

Figure 3.9: Ambiguous $\text{SFI}_{\gamma=1}$ data and artifacts of background estimation. Panel (a) shows early dust plumes at 7.30 a.m. GMT (blue arrows) and optically similar appearing dust free areas (white arrows). In Panel (b) the black arrows point out artefacts of the background estimation procedure.

of the object. For instance, the infrared radiation emitted by the earth's surface varies with the respective surface temperature which itself depends on factors like incidence of the sun's radiation (and hence day time) and vegetation. While barren soil heats up and cools down relatively fast with insolation increase and decrease, vegetated areas remain comparably stable in their heat (and thus infrared) emission due to respective regulatory mechanisms of plants.

A satellite product that is particularly expressive in terms of surface vegetation and hence general infrared radiation levels as well as day time variability surface emissivities at $8.3\mu m$. In the approach to dust detection at hand we employ respective data generated according to the algorithm of Seemann et al. (2008). The algorithm derives monthly averages of surface emissivity at SEVIRI's spatial resolution from the MODIS instrument aboard the non-stationary NASA platform Terra. An additional advantage of choosing this particular wavelength is that, from our experience, it seems comparably less disturbed by satellite swath artifacts. Figure 3.10 provides a visual comparison of the employed emissivity estimate, the corresponding $\text{SFI}_{\gamma=1}$ as well as visual spectra data as shown by the Google Earth software. Clearly, the former picks up the vegetative range perceivable from the blue to green color range of the latter. From here on, these emissivity estimates are denoted by $E(x, y, t)$ for coordinates (x, y) within the image domain Ω and time t.

Given the described emissivity estimates we let E supersede the anomaly indicating term $\mathbf{A} - \mathbf{I}$ in the same LSM BHM described in Section 3.3.1. The linear predictor



(a) (b) (c)

Figure 3.10: MODIS emissivity estimates at 8.3 μm. Panel 3.10(a) shows the average emissivity of the earth's surface at the west coast of northern Africa during January 2010. To allow for a visual comparison with $\text{SFI}_{\gamma=1}$ data (Panel 3.10(b), January 14) the emissivity values are colored such as to range from 0.6 (green) to 1 (blue). Panel 3.10(c) shows the same region using a visual spectra visualization by the Google Earth software and indicates vegetative properties.

is then of the form

$$\eta(x,y,t) \quad = \quad \sum_{i=1}^{3} g_i(I_{ixyt}, E_{xyt}), \tag{3.15}$$

where each of the new functionals g_i is again modeled using a CAR with log Gamma distributed hyper parameter priors. As opposed to our previous studies we employ the SFI representation of the SEVIRI data instead of $\text{SFI}_{\gamma=1}$, that is

$$\mathbf{I} = (I_{1xyt}, I_{2xyt}, I_{3xyt}) = \text{SFI}(\Delta T_{BR}, \Delta T_{BG}, BT_{10.8}).$$

This is due to two reasons. Firstly, in the following we will compare our approach with the thresholding technique by Ashpole and Washington (2012) introduced in Section 3.1.2, which is based on the SFI representation as well. Secondly, the aforementioned simulation study presented by Brindley et al. (2012) (see Section 3.1.2) yields a connection between SFI and the AOD at $\tau_{10} = 10\mu m$. It is therefore sensible to evaluate our method in the SFI space to facilitate future work on comparing or fitting the dust predictor to AOD products or simulated data.

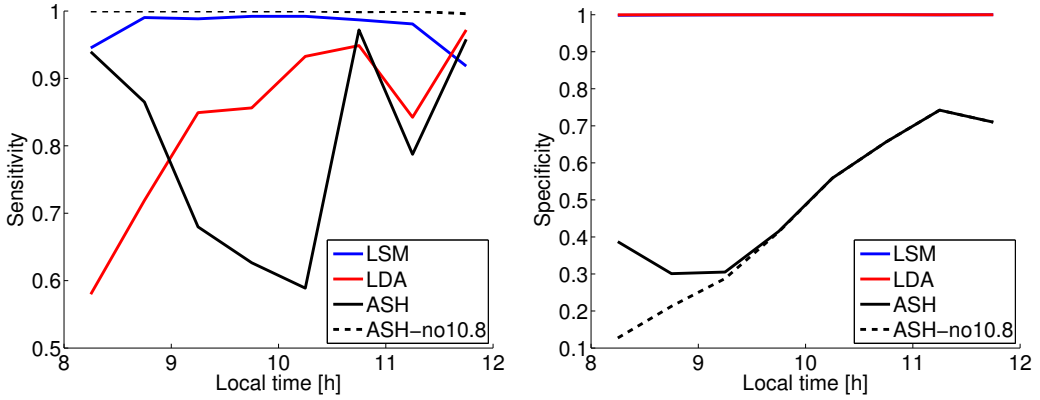
3.4.2 Experiments & Results

The basis of the following analysis is a SEVIRI data set spanning January 10–26, 2010, a period with several smaller and large scale dust events. By visual inspection we performed an extensive labeling of dusty and pristine regions. The data set contains extensive samples of pristine and dust contaminated areas across northern Africa and the Arabic peninsula at local times between 8 a.m. and noon.

With these samples at hand we conducted a two-fold cross validation study. This procedure splits the samples into two randomly chosen disjoint sets, one of which serves to perform inference on the model via the given labels. The samples of the other (test) set are employed to infer their dust probability and compare it to the manually declared labels. These samples were flagged as dusty whenever the respective probability was above 0.5. In a second run the roles of the sets are exchanged and subsequently the performance results of the runs are averaged. The SEVIRI signal changes strongly with the relative position of the sun and dust plume genesis often predominantly occurs during the forenoon. Thus, in order to assess the prediction performance as a function of the local time of the pixel, samples of the respective test set were grouped according to their time stamp. As a last step, within group sensitivity and specificity was computed. We compared the performance of four methods for estimating the probability of dust, the LSM approach, a simple linear discriminant analysis (LDA), and two thresholding approaches introduced by Ashpole and Washington (2012). In case of LDA and LSM a pixel is classified as dusty if the probability of dust is greater than 0.5, and as pristine otherwise. The first approach of Ashpole and Washington (ASH-no10.8) determines a pixel to be dusty if Equations 3.1 and 3.2 presented in Section 3.1.2 hold. For the second method (ASH), also Equation (3.3) is required to hold. Figure 3.11 shows the percentage of correctly classified clear pixels (left panel) and those containing dust (right panel), stratified by the time of day of the image.

From Figure 3.11 we draw several interesting conclusions. First, we see that the two thresholding approaches perform poorly in correctly classifying clear, or pristine, regions. Even the more involved ASH leads only to slight improvements. By contrast, the simpler ASH-no10.8 thresholding approach performs essentially perfectly at classifying clear regions while the additional threshold of ASH significantly decreases the fraction of correctly recognized dusty samples. By contrast, the LDA perfectly classifies pristine areas, but performs poorly during the early hours (between 8 am and 10 am) at classifying dusty pixels. Finally, the LSM method considerably improves on LDA for dusty pixels and achieves nearly perfect classification in both situations throughout the entire time frame. These results extend those found in Bachl et al. (2013a) and justify our use of the LSM emissivity modeling approach in (3.15) on these data.

Figure 3.12 provides some indication of why LSM improves over LDA and thresholding. In this figure, the left column shows pixels labeled as clear, or pristine, while the



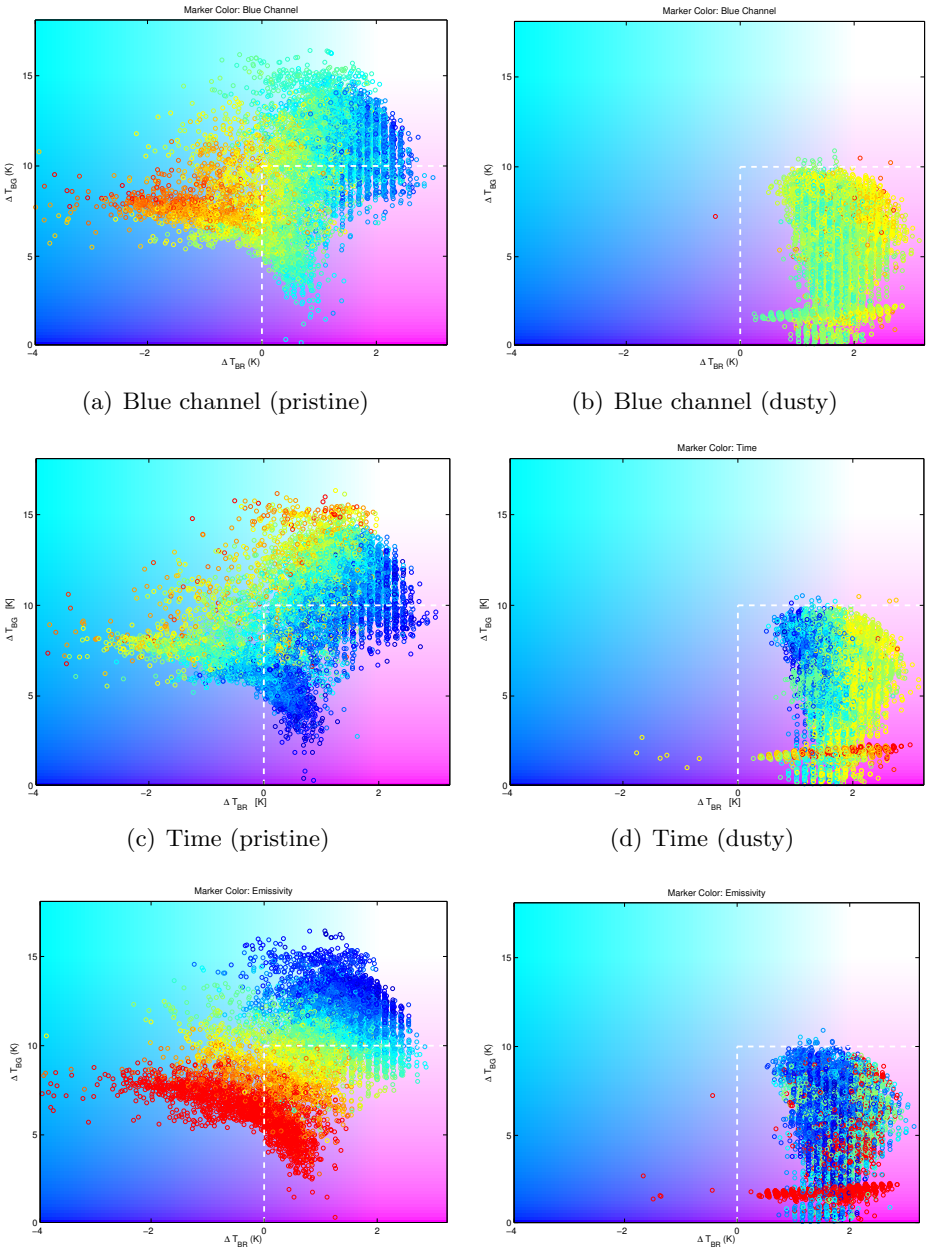
(a) Dusty pixels

(b) Pristine pixels

Figure 3.11: Cross validation results for pixel-wise dust detection under the LSM emissivity approach (blue), linear discriminant analysis (red) and the two thresholding methods of Ashpole and Washington (2012) (black). The plots show the percentage of correctly classified (a) dusty pixels and (b) pristine pixels, stratified by the hour of the day.

right hand column pertains to dust-filled pixels. In each figure, points are placed relative to their green channel intensity (x-axis) and red channel intensity (y-axis). Dotted lines show the thresholding cut-offs of Ashpole and Washington (2012). From the dotted lines, we immediately see why the thresholding approach performs poorly at classifying clear pixels, a large portion are inside the threshold.

Figure 3.12 also demonstrates why LDA alone performs poorly in the early hours. In the first row points are colored according to the local time at which the data was collected with earlier time points shown in blue. As we can see, the red and green channel intensities for both dusty and clear points are initially very similar, while subsequently the intensities begin to diverge. Since the LDA method classifies the data based on these intensities only, it struggles in the early hours while it improves significantly as the day progresses. The emissivity information in the data is displayed in the bottom row of Figure 3.12. For clear pixels there is a strong relationship between green and red channel intensity and emissivity levels. By contrast, for dusty pixels, the emissivity has no relation to channel intensity since strong dust events completely block 8.3μm radiation. In combining this information with channel intensity in the LSM approach, we thus achieve an improved classification in the early-morning data.



(a) Blue channel (pristine)

(b) Blue channel (dusty)

(c) Time (pristine)

(d) Time (dusty)

(e) Emissivitiy (pristine)

(f) Emissivity (dusty)

Figure 3.12: Green channel intensity (x-axis) versus red channel intensity (y-axis) of the labeled training data. Left column: pixels labeled as pristine; right column: pixels labeled as dusty; top row: points are colored by local time of the day, from blue (early) to red (later); bottom row: points are colored by emissivity, from blue (low) to red (high). The white dashed lines indicate the "no10.8" thresholding of Ashpole and Washington (2012). As the entire data set is very large, each plot shows a random subsample of the full data set.

3.4.3 Conclusion

LSM with surface emissivities cumulates several steps of developing a Bayesian framework for dust detection. With the presented cross validation study it becomes clear that this approach is significantly better in extracting dust evidence from SEVIRI data than simple thresholding schemed and LDA. In particular, the method is capable to indicate dust presence even during the early hours of the day where the radiative conditions hamper even visual assessment of the data. Although this was not tested within the study at hand, it is likely that due to the symmetry of the daily sun cycle a similar statement holds for late afternoon observations. In summary, the model for η appears to work quite well in our current data but considerable work remains from application and methodological perspectives. It could be extended in several manners. The most useful of these extensions would be to make the estimates of η depend not just on emissivity and image intensity, but to also include spatial and temporal dependence on neighboring estimates. In practice this appeared to be unnecessary in our current approach and the computational challenges to such estimation proved challenging. However, as the performance of the INLA software continually improves, such developments may become helpful. Another worthwhile extension would be to take the local time or other covariates such as satellite viewing angle of a particular pixel location into account. In particular if the dust analysis is extended from the forenoon to a whole day (e.g. night time) the former might be a critical feature to prevent a degradation of detection performance.

4 Dust Transport

The previous chapter elucidated the development of a framework for estimation of dust aerosol presence from remote sensing data. Once the respective linear predictor $\eta(x, y, t)$ of dust probability is determined a multitude of applications motivate to model their dynamics in both space and time. For instance, one would be interested in forecasting the movement of a dust plume or determine a connection between wind field measurements and the behavior of a dust storm. Statistically capturing the transport process of dust, i.e. its *flow field*, through the atmosphere and facilitating these applications is the central objective of this chapter.

The physical laws of dust motion are examined within the discipline of *fluid dynamics*. These form the first part of the preliminaries Section 4.1 and provide the reader with a physical interpretation of the methods developed throughout this Chapter. Within the scientific field of image analysis *variational inference* is one of the most prominent approaches to model and solve problems like motion estimation and will receive attention in Section 4.1.2. An example for such an approach is the optical flow method of Horn and Schunck (1981) introduced in Section 4.1.3 and from here on abbreviated by HS. It was one of the first to yield motion estimation from image data and constitutes a fundamental building block of the methodology developed within this thesis. The preliminaries section of this chapter is concluded by Section 4.1.4, a brief overview on a very recent development in computational statistics in terms of connections between GMRFs (see Section 2.6) and stochastic partial differential equations (SPDEs).

Subsequently, three thematic blocks are covered. In Section 4.2 a link between the HS method and BHMs (see Section 2.4) is established. This way, the transport process of the dust is interpreted as a latent GMRF and posterior inference as well as hyper parameter integration via integrated nested Laplace approximation (INLA) (see Section 2.7) is straight forward. First encouraging results on dust transport and source detection are shown, which are taken from Bachl and Garbe (2012).

A main motivation to develop the LSM method for dust detection was the suppression of noise entering the predictor in pristine regions. Section 4.3 offers a respective case study that is taken from Bachl et al. (2012). It is shown that dust flow derived from a LSM predictor is less affected by noise than flow based on predictors estimated by linearly projecting methods.

The Horn and Schunck method implies that the flow of the moving substance is *incompressible* and thus divergence free, an assumption that is violated in case of remotely sensed aerosols. Section 4.2 elucidates how to specify the previously developed Bayesian motion estimation framework such as to satisfy the *continuity equation*. Thereby physical effects like satellite column projection and *compressible flow* become intrinsic parts of the model. The importance of this step by is demonstrated by a multitude of experiments taken from Bachl et al. (2013a) and Bachl et al. (2013b).

The flow estimation technique developed within this thesis is dedicated to combine methodology from physics, mathematical image processing and Bayesian statistics. Section 4.5 elaborates on a fruitful combination of these disciplines with respect to theoretical properties of flow estimation. Using variational techniques from image processing we present a proof that guarantees existence and uniqueness of compressible flow over a continuous domain under mild restrictions in terms of an underlying Sobolev space. As a consequence, we are able to show that the corresponding discretized version leads to a posterior with a proper precision matrix. The section is concluded with very recent work leaning towards estimation of dust presence given the respective flow and thus, in the long run, a joint probabilistic estimation technique. Here, we point out first variational steps to express the dust predictor as a solution to a class of SPDEs for which recent results of Lindgren et al. (2011) provide analytical results in terms of GMRF approximations and the applicability of the INLA technique.

4.1 Preliminaries

4.1.1 Fluid Dynamics

Consider a fluid moving with velocity $\mathbf{v}$ and carrying a material density ρ inside a spatial volume S. The mass of the material is then a volume integral of the density: $m = \iiint_S \rho \, \mathrm{d}S$. By the law of conservation, the change of mass in this volume over time equals the mass flux $\mathbf{J} = \rho\mathbf{v}$ across the boundaries $\bar{S}$ of the volume. Moreover, the divergence theorem allows to express this boundary flux as an integral of the mass flux divergence over the volume S:

$$\frac{\partial m}{\partial t} = - \oint_{\bar{S}} \mathbf{J} \cdot \mathrm{d}\bar{S} = - \iiint (\nabla \cdot \mathbf{J}) \, \mathrm{d}S. \tag{4.1}$$

If we now focus on an infinitesimal volume this gives a connection of the density's temporal differential to the mass flux,

$$\frac{\partial m}{\partial t} = \frac{\partial}{\partial t} \iiint_V \rho \, \mathrm{d}V = \iiint_V \frac{\partial \rho}{\partial t} \, \mathrm{d}V = - \iiint_V (\nabla \cdot \mathbf{J}) \, \mathrm{d}V, \tag{4.2}$$

from which we can see that

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot \mathbf{J} = -\nabla \cdot (\rho \mathbf{v}) = -(\nabla \rho \cdot \mathbf{v} + \rho(\nabla \cdot \mathbf{v})) \tag{4.3}$$

and hence

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = \frac{\partial \rho}{\partial t} + \mathbf{v} \cdot \nabla \rho + \rho(\nabla \cdot \mathbf{v}) = 0, \tag{4.4}$$

which is called the *continuity equation*. So far we considered a control volume with a fixed position in space and time, i.e. the Eulerian perspective. Here, the temporal derivative of the mass/density must not vanish as material is allowed to flow through the control volume. An important consequence of Equation (4.3) arises when taking the Lagrangian perspective on the motion of the material. If we assume that the control volume follows a trajectory $[x, y, z](t)$ we can characterize the change of the density along this path via its total derivative

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = \frac{\partial \rho}{\partial t} + \frac{\partial \rho}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial \rho}{\partial y}\frac{\partial y}{\partial t} + \frac{\partial \rho}{\partial z}\frac{\partial z}{\partial t}. \tag{4.5}$$

Substitution of the trajectory by the given velocity field and using the continuity equation then gives

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = \frac{\partial \rho}{\partial t} + \nabla \rho \cdot \mathbf{v} = -\rho(\nabla \cdot \mathbf{v}). \tag{4.6}$$

This means that if the control volume moves according to the velocity inherent to the flux and the density is non-zero, the change of the latter inside the volume depends on the divergence $\nabla \cdot \mathbf{v}$ of the flux.

The middle term of Equation (4.6), i.e. the material derivative, describes how the density of the material changes if the control volume follows the flow of the transporting medium. Physically, flows are characterized by their compressibility. *Isochoric (incompressible) flows* require that the material derivative of the density vanishes, i.e. that for non-zero ρ the term divergence of the flow equates to zero. In this case the continuity equation reduces to

$$\frac{\partial \rho}{\partial t} + \nabla \rho \cdot \mathbf{v} = 0. \tag{4.7}$$

4.1.2 Variational Inference

In image processing mathematical objects of interest are often characterized in terms of energy functionals. The object of interest can for instance be a real function $\mathbf{v} : \Omega \to \mathbb{R}$ over the image domain Ω that represents a continuous extension of the discrete image given by values at pixel locations. As we will see later on, $\mathbf{v}$ can as well yield a mapping of the image data to multiple local properties (i.e. $\mathbb{R}^n$) like the motion field that underlies changes in a sequence of images.

Similar to an unnormalized probability distribution the functional describes several requirements imposed on $\mathbf{v}$ together with a penalization in terms of deviations. In case of continuous problems a convenient notation is to instantiate a functional $\mathcal{J}$ as an integral over the image domain Ω with respect to a local requirement L that depends on the sought solution $\mathbf{v}$ and corresponding differentials like the Jacobian $J_{\mathbf{v}}$:

$$\mathcal{J}(\mathbf{v}) = \int_{\Omega} L(\mathbf{x}, \mathbf{v}, J_{\mathbf{v}}) \, \mathrm{d}\mathbf{x}.$$

The function $\hat{\mathbf{v}}$ that, so to say, optimally meets the requirements on average (or, equivalently, minimizes the average local penalization) is then characterized as to minimize the functional. That is,

$$\hat{\mathbf{v}} = \mathrm{argmax}_{\mathrm{v}} \, \mathcal{J}(\mathbf{v}).$$

These problems are then solved using the fact that the functional derivative with respect to to $\hat{\mathbf{v}}$ equates to zero. Hence, the Euler-Lagrange equations can be employed such that $\hat{\mathbf{v}}$ has to solve

$$\frac{\partial L}{\partial \hat{v}_k} - \sum_i \frac{\mathrm{d}}{\mathrm{d}x_i} \frac{\partial L}{\partial \hat{v}_{k,x_i}} = 0 \quad \forall k, \tag{4.8}$$

where the total differential reduces to partial derivatives if the spatio-temporal coordinates x_i are assumed to be independent. Typically, these equations are solved on a gridded discretization, i.e. at the pixel locations of an image.

Example: Image Smoothing

Consider a noisy image $\bar{g}$ as a function $\bar{g} : \Omega \subset \mathbb{R}^2 \to \mathbb{R}$ over the domain Ω. We can think of the original image as to obey two criteria. Firstly, its point-wise squared distance to the noisy image is supposed to be small. Secondly, one might assume that the original image is relatively smooth and thus its gradients are relatively small. Combining these criteria and integrating them over the image domain one obtains the functional

$$\mathcal{J}(g) = \int_{\Omega} (g - \bar{g})^2 + ||\nabla g||^2. \tag{4.9}$$

Following the approach of variational inference the partial derivatives are then readily obtained and result in the Euler-Lagrange equation

$$\bar{g} = (1 - \Delta) g. \tag{4.10}$$

From this one can get an idea of how the constituents of the functional $\mathcal{J}$ enter a partial differential equation expressing the problem:

1. The smoothness assumption with respect to the gradient generates a Laplace operator Δ applied to the image g and hence represents the corresponding diffusion. In the further reading of Section 4.1.4 one should note that this exactly gives the Laplacian part of Equation (4.25) (if $\alpha = 2$) and Equation (4.26) since $\Delta g = \nabla^T I \nabla g = \nabla^T \nabla g$.

2. The quadratic image fit constitutes the inhomogeneous part of the partial differential equation (PDE) (via $\tilde{g}$) as well as the direct occurrence of g in the latter (see the corresponding factor 1 inside the parentheses of the SPDEs 4.25 (for $\kappa = 1, \alpha = 2$) and Equation (4.26).

3. Another way of looking at the latter connection is that for $\tilde{g} \equiv 0$ a white (uniform) Gaussian prior is generated.

We are concerned with a particular class of functionals for which existence and uniqueness of the minimizing argument $\mathbf{v}$ can be guaranteed under certain conditions. For this purpose let from here on $(V, |\cdot|_V)$ denote a Hilbert space for which we wish to determine $\mathbf{v} \in V$ such as to minimize a functional $\mathcal{J}$. Assume furthermore that $\mathcal{J}$ attains the form

$$\mathcal{J}(\mathbf{v}) = \frac{1}{2} a(\mathbf{v}, \mathbf{v}) - f(\mathbf{v}) + c, \tag{4.11}$$

where $a(\cdot, \cdot) : V \times V \to \mathbb{R}$ is a continuous bilinear form, $f(\cdot) : V \to \mathbb{R}$ is a continuous linear form and $c \in \mathbb{R}$ is a constant. Given these conditions the lemma of Lax-Milgram (see for example Hackbusch (2005)) provides a convenient set of conditions for the uniqueness and existence of the minimizer solving Equation (4.1.2).

Lemma 3 (Lax-Milgram)**.** *Let $(V, |\cdot|_V)$ denote a complete Hilbert space and let $\mathcal{J} : V \to \mathbb{R}$ of the form given in Equation* (4.11)*. Let $a(\cdot, \cdot)$ symmetric as well as V-elliptic, that is,*

$$a(\mathbf{v}, \mathbf{v}) \geq C |\mathbf{v}|_V^2 \quad \forall \mathbf{v} \in V, C > 0. \tag{4.12}$$

Then $\mathcal{J}$ attains its unique minimum at the solution $\mathbf{u} \in V$ to the following problem:

$$a(\mathbf{u}, \mathbf{v}) = f(\mathbf{v}) \qquad \forall \mathbf{v} \in V. \tag{4.13}$$

Lemma 3 and in particular Equation (4.13) play an important role in the discretization of the problem in applied contexts. For suitably chosen basis functions of V it can be shown that the *Ritz method* of discretizing Equation (4.13) becomes a finite-element method. We now briefly sketch its main idea, for details the reader is referred to Hackbusch (2005) and Schnörr (1991).

The Ritz method uses a finite-dimensional subspace $V_N = \text{span}\{\phi_1, \ldots, \phi_N\} \subset V$ to approximate the solution to Eq. 4.13. A valid basis $\{\phi_1, \ldots, \phi_N\}$ in case of a square image domain would, for instance, be a set of square indicator functions

corresponding to area covered by a pixel in the image. Now let $\hat{\mathbf{v}} = (\hat{v}_1, \ldots, \hat{v}_N)$ denote the coefficient vector that generates an element $\mathbf{v}$ of V_N via the mapping

$$P : \mathbb{R}^n \to V_N, \quad \mathbf{v} = P\hat{\mathbf{v}} := \sum_{j=1}^{N} \hat{v}_j \phi_j \tag{4.14}$$

and let $L_{ij} = a(\phi_i, \phi_j)$ as well as $b_j = f(\phi_j)$. Then the left and right hand side of Equation (4.13) read as

$$a(\mathbf{u}, \mathbf{v}) = a(P\hat{\mathbf{u}}, \hat{\mathbf{v}}) = \hat{\mathbf{u}}^T L \hat{\mathbf{v}}$$

and

$$f(\mathbf{v}) = f(P\hat{\mathbf{v}}) = \mathbf{b}^T \hat{\mathbf{v}}.$$

Now, note that since a is symmetric and V-elliptic it is easy to see that L is SPD. With that, the Ritz-solution $\hat{\mathbf{u}}$ to our problem is at the same time the unique solution to the system

$$L\hat{\mathbf{u}} = \mathbf{b}. \tag{4.15}$$

Hence, once a suitable basis is selected, a linear system is readily set up and solved for, yielding a discretized solution of the functional one attempts to minimize.

4.1.3 The Horn & Schunck Method for Optical Flow

Most motion estimation techniques are based on the assumption that there is a photometric or geometric quantity of the image sequence that is preserved spatially or temporally. One of the first approaches to this subject was was brought up by Horn and Schunck (1981). The main idea is that along the trajectory of a moving object the intensity that this object contributes to the image intensity stays constant, which is expressed by the co called brightness constancy equation (BCE). For a given triplet (x, y, t), suppose that $\eta(x, y, t) = k$. The BCE then stipulates that there is a path $(x(r), y(r))$ in $\Omega \times [0, T]$, such that for all $r \in [0, T] \subset \mathbb{R}$

$$\eta(x(r), y(r), r) = k. \tag{4.16}$$

From this, two analytical derivations lead to the method applied by Horn and Schunck. First, it is easy to see that the BCE can be understood as a constraint on two consecutive images or spatial predictions $\eta(x, y, t)$ and $\eta(x, y, t + \Delta t)$, i.e.

$$\eta(x, y, t) = \eta(x + \Delta x, y + \Delta y, t + \Delta t). \tag{4.17}$$

A Taylor series expansion then leads to

$$\eta(x + \Delta x, y + \Delta y, t + \Delta t) = \eta(x, y, t) + \frac{\partial \eta}{\partial x}\Delta x + \frac{\partial \eta}{\partial y}\Delta y \frac{\partial \eta}{\partial t}\Delta t + h, \tag{4.18}$$

where h are the higher order terms. Now let $\Delta x/\Delta t = u(x,y)$ and $\Delta y/\Delta t = v(x,y)$ denote spatially dependent change in coordinates, i.e., the motion field. Then, dropping the higher order terms, Equation (4.18) reduces to

$$\eta(x+\Delta x, y+\Delta y, t+\Delta t) \;=\; \eta(x,y,t) + \frac{\partial \eta}{\partial x}u + \frac{\partial \eta}{\partial y}v + \frac{\partial \eta}{\partial t}, \tag{4.19}$$

which together with the BCE gives

$$0 = \frac{\partial \eta}{\partial x}u + \frac{\partial \eta}{\partial y}v + \frac{\partial \eta}{\partial t}. \tag{4.20}$$

Equivalently to this approach one can employ the total derivative along the trajectory $(x(r), y(r))$. Assuming no change in image intensity and no higher order dependencies of x and y (i.e. $\mathrm{d}x/\mathrm{d}t = \partial x/\partial t$ and $\mathrm{d}y/\mathrm{d}t = \partial y/\partial t$) it then holds that

$$\begin{aligned} 0 = \frac{\mathrm{d}}{\mathrm{d}t}\eta &= \frac{\partial}{\partial t}\eta + \frac{\mathrm{d}x}{\mathrm{d}t}\frac{\partial}{\partial x}\eta + \frac{\mathrm{d}y}{\mathrm{d}t}\frac{\partial}{\partial y}\eta \\ &= \eta_t + \frac{\mathrm{d}x}{\mathrm{d}t}\eta_x + \frac{\mathrm{d}y}{\mathrm{d}t}\eta_y \\ &\approx \eta_t + u\,\eta_x + v\,\eta_y, \end{aligned}$$

where the dependence on (x,y,t) has been dropped. Most importantly, this is expression is equivalent to the continuity equation for incompressible fluid flow (Equation (4.7)).

With two unknowns this equation is under-determined, an issue known as the aperture problem (see Figure 4.1). As with many other approaches, the HS optical flow therefore imposes an additional constraint. In order to maintain physical plausibility and to propagate information into image regions with ambiguous gradient properties, non-smoothness of the flow is penalized via the Euclidean norm of the gradient. The final optical flow is then defined as the minimizer of the squared deviations of the BCE fit plus the smoothness term L integrated over the image domain Ω. That is,

$$(u,v)(\alpha) = \mathrm{argmin}_{u,v}\mathcal{J}_{\mathrm{HS}}(\alpha),$$

where α is a regularization parameter and

$$\mathcal{J}_{\mathrm{HS}}(\mathbf{v}) = \int_\Omega L_{\mathrm{HS}}(\mathbf{v}) + \alpha^2 L_{\mathrm{S}}(\mathbf{v}) \tag{4.21}$$

with a data term

$$L_{\mathrm{HS}}(\mathbf{v}) = (\eta_t + u\,\eta_x + v\,\eta_y)^2, \tag{4.22}$$

and smoothness term

$$L_{\mathrm{S}}(\mathbf{v}) = (|\nabla v_1|^2 + |\nabla v_2|^2). \tag{4.23}$$

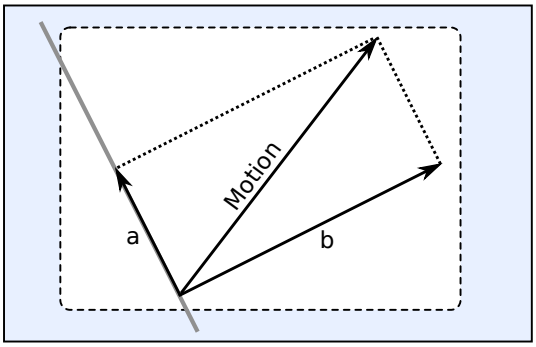


Figure 4.1: The aperture problem. Consider a scenery (outer box) and a moving object (grey line). Having only local knowledge can be seen as to perceive the scenery through an aperture (dashed box). The object's motion component a perpendicular to its gradient b is not visible through the aperture.

The variational approach to finding the minimizer of the HS functional employs the respective Euler-Lagrange equations,

$$\begin{aligned} \eta_x(\mathbf{v}\cdot\nabla\eta+\eta_t)-\alpha^2\Delta v_1 &= 0, \\ \eta_y(\mathbf{v}\cdot\nabla\eta+\eta_t)-\alpha^2\Delta v_2 &= 0. \end{aligned}$$

This system of equations can readily be solved by an application of Gauss-Seidel iterations

$$\begin{aligned} u^{n+1} &= \bar{u}^n-\eta_x\frac{\eta_x\bar{u}^n+\eta_y\bar{v}^n+\eta_t}{\alpha^2+\eta_x^2+\eta_y^2}, \\ v^{n+1} &= \bar{v}^n-\eta_y\frac{\eta_x\bar{u}^n+\eta_y\bar{v}^n+\eta_t}{\alpha^2+\eta_x^2+\eta_y^2}, \end{aligned}$$

with local averages $\bar{u}$ and $\bar{v}$ of u and v, respectively. Most importantly, however, Schnörr (1991) shows that the existence and uniqueness of the solution with respect to the flow can be guaranteed in case of the assumption that the flow field comes from a mildly restricted Sobolev space.

Theorem 7 (Schnörr, 1991)**.** *Let* $H^1(\Omega)=\{u\in L^2(\Omega): D^\psi u\in L^2(\Omega) \text{ for } |\psi|\leq 1\}$ *denote a Sobolev space and let* $V=\{\mathbf{u}=(u_1,u_2)^T\in H^1(\Omega)\times H^1(\Omega)\}$*. If* $\alpha>0$*,* $\eta_x,\eta_y\in L^\infty(\Omega)$*,* η_x *and* η_y *are linearly independent as elements of* $L^2(\Omega)$ *then the functional* $\mathcal{J}_{\mathrm{HS}}: V\to\mathbb{R}$ *attains a unique minimum that depends continuously on the image data* η*.*

4.1.4 Stochastic Partial Differential Equations and GMRFs

In a number of scientific fields (see, e.g. Guttorp and Gneiting (2006)) the phenomenon under observation naturally follows a probabilistic description involving the Matérn

covariance function

$$r(\mathbf{x},\mathbf{y}) = \frac{\sigma^2}{\Gamma(\nu)2^{v-1}}(\kappa||\mathbf{x}-\mathbf{y}||)^{\nu}K_{\nu}(||\mathbf{x}-\mathbf{y}||), \tag{4.24}$$

where $||\mathbf{x}-\mathbf{y}||$ denotes the Euclidian distance for locations $\mathbf{x},\mathbf{y} \in \mathbb{R}^d$. As shown by Whittle (1954, 1963), this is not the only way to characterize these phenomena and there exist close connections to stochastic partial differential equations (SPDEs). Most importantly, if $\mathcal{W}$ is a Gaussian field with (spatial) Gaussian white noise and unit variance, the solution $g(\mathbf{x})$ of

$$(\kappa^2-\Delta)^{\alpha/2}g(\mathbf{x}) = \mathcal{W}(\mathbf{x}), \qquad \mathbf{x}\in\mathbb{R}^d, \quad \alpha=\nu+d/2, \quad \kappa>0, \quad \nu>0 \tag{4.25}$$

follows the Matérn covariance function. Recently, further connections between Matérn fields and SPDEs have been discovered. While it is been known (Besag, 1981) for a while that the solution to Equation (4.25) can be approximated by an autoregressive GMRF on a unit distance regular grid for $\kappa^2 = a-4, \nu = 0$ and $\sigma^2 = q/(4\pi)$, Lindgren et al. (2011) were able to extend this result significantly. As a comprehensive discussion of these findings is beyond the scope of this paper we will restrict ourselves to the following summarizing statements. Firstly, stochastic weak solutions of the SPDE in Equation (4.25) can be approximated by a particular GMRF representation on a triangulated lattice and the computational cost of constructing this GMRF is $\mathcal{O}(n)$, where n is the number of triangulation points. Secondly, the very same approach can be employed to compute approximative solutions to SPDEs defined on regular manifolds like spheres and by introducing space-varying parameters non-stationary and non-isotropic fields are modeled in a coherent fashion. The latter finding is particularly important for environmental sciences and remote sensing. Even on a simple $\mathbb{S}^2$ sphere it is not possible to define a valid a Matérn field that acts on great circle distance correlations (see again Lindgren et al. (2011) for details). A model like the SPDE that naturally extends the Matérn family to such manifolds is therefore highly appreciated to set forth interpretability and analytical consistency.

The very same publications also point out an SPDE interpretation of the image warping method of Sampson and Guttorp (1992), which itself is closely related to the method of optical flow we will elaborate on later. Here, a mapping $f : \tilde{\mathbf{x}} \in \tilde{\Omega} \rightarrow \mathbf{x} \in \Omega$ from an deformed space $\tilde{\Omega}$ onto a undeformed space Ω acts as a transformation of a stationary SPDE

$$(1-\tilde{\nabla}^T\mathbf{m}-\tilde{\nabla}^T\tilde{\nabla})\tilde{g}(\tilde{\mathbf{x}}) = \tilde{\sigma}\mathcal{W}(\tilde{\mathbf{x}}) \tag{4.26}$$

into a non-stationary and non-isotropic formulation

$$\frac{1}{\det(F)}\Big(1-\nabla^T F\mathbf{m}-\det(F)\nabla^T\frac{FF^T}{\det(F)}\nabla\Big)g(\mathbf{x}) = \frac{\tilde{\sigma}}{\det(F)^{1/2}}\mathcal{W}(\mathbf{x}) \tag{4.27}$$

via a change of variables (F is the Jacobian of the deformation function). This generalizes the method of Sampson and Guttorp insofar as their approach employs a Matérn field and allows to ignore first order derivatives by letting $\mathbf{m} = 0$.

4.2 A BHM for Incompressible Optical Flow

Rheology, the study of the flow of liquid matter, and the motion estimation of quasi-rigid bodies has been an active research field of image processing and computer vision during the last two decades. With respect to image analysis in experimental fluid dynamics these efforts led to an increasing expertise in correlation-based particle image velocimetry methods and variational approaches to the problem (see Heitz et al. (2010) for a review on this topic). Similar frameworks have been developed in computational statistics due to the increasing interest in modeling spatio-temporal processes for environmental science applications, e.g. ozone and precipitation interpolation and forecasting (Glasbey and Mardia, 1998). In particular, methods based on the perspective of image deformation (warping), e.g. by Aberg et al. (2005), are related to our work.

However, to the best of our knowledge, the connection between probabilistic and variational approaches is reflected only by a few publications. Simoncelli et al. (1991) point out the distributional aspects of the Horn & Schunck method for optical flow (HS). A maximum-a-posteriori approach to the free parameters of this method was illustrated by Krajsek and Mester (2006b). Krajsek and Mester (2006a) further show the limit-equivalence of the variational solution of the HS functional to the mode of a normal distribution defined via the maximum entropy principle with respect to observations at discretized locations.

In what follows we first describe how the HS method can be expressed as inference in a BHM (Section 4.2.1). The underlying idea was first published by Bachl and Garbe (2012) together with an encouraging experiment in terms of dust flow estimation and spatial indication of source regions. The experiment is presented in Section 4.2.2 and respective conclusions are drawn in Section 4.2.3. Note that, for convenience, we will refer to the flow determined via the HS method (and our BHM as well) as *incompressible*. This is intended to reflect the respective intrinsic assumption of the HS method. As deviations from this assumptions are allowed within the HS formulation, the resulting flow might, however, not be incompressible.

4.2.1 Methods

As before, fix (x, y) for coordinates within the image domain Ω. We then aim to determine the vector field $\mathbf{w}(x, y, t) = (u(x, y, t), v(x, y, t))$, where $u(x, y, t)$ and $v(x, y, t)$ are the instantaneous change in $\eta(x, y, t)$ in the vertical and horizontal directions. Now recall that the general idea of the Horn and Schunck approach is expressed by the brightness constancy equation, which ultimately leads to the equation

$$\eta_x u + \eta_y v + \eta_t = 0. \tag{4.28}$$

The method then imposes a squared penalization upon deviations of the left hand side from zero integrated over the image domain (see the data term L_{HS} in Equation (4.22)). In the discrete sense the BCE error term is then equivalent to an interpretation of the image gradients as an observational system of the latent flow variables u and v with additive Gaussian noise. That is,

$$u\,\eta_x + v\,\eta_y = -\eta_t + \epsilon, \qquad \epsilon \sim \mathcal{N}(0, \sigma^2).$$

It follows that the partial derivatives of $\eta(x, y, t)$ define a Gaussian likelihood $p(\nabla\eta|\mathbf{u}, \mathbf{v})$ for the discretized optical flow (see Section 2.5 for details on the respective GLMs with Gaussian observations). The regularization term is discretized by approximating the integral over the image domain Ω with the Riemann sum over a regular grid $\mathcal{G} \subset \Omega$, i.e.

$$\int_\Omega \alpha^2 |\nabla u|^2 = \alpha^2 \int_\Omega u_x^2 + u_y^2 \approx \alpha^2 \sum_{(i,j)\in\mathcal{G}} u_x^2(i,j) + u_y^2(i,j)$$

in case of the horizontal flow gradient ∇u and equivalently for ∇v. The partial derivatives are approximated by horizontal and vertical differences, i.e. $u_x(i,j) = u_{ij} - u_{i-1,j}$ and $u_y(i,j) = u_{ij} - u_{i,j-1}$. By summing up both over all grid points it then follows that the regularization part of L_{HS} related to ∇u reduces to

$$\int_\Omega \alpha^2 |\nabla u|^2 \approx \alpha^2 \sum_{s_1 \sim s_2} (u_{s_1} - u_{s_2})^2,$$

where $s_1 \sim s_2$ denotes the set of all unordered grid neighbors s_1 and s_2 (for details see Rue and Held (2005), Section 3.2.2). This formulation is analytically identical to the log-density of a CAR GMRF, illustrating the equivalence of the estimation of HS optical flow and Bayesian modeling of spatially dependent systems (Besag, 1974). Thus, the smoothness part of the HS functional defines GMRF priors $p(u) = \mathcal{N}(\mathbf{0}, Q_u)$ and $p(v) = \mathcal{N}(\mathbf{0}, Q_v)$ for the latent flow fields if the precision matrices are defined via

$$Q_{ij}(\alpha) = \alpha^2 \begin{cases} n_i, & i = j \\ -1, & i \sim j \\ 0, & \text{otherwise} \end{cases}$$

where n_i is the number of neighbors on the grid. This formulation also clarifies the role of the smoothness parameter α as a hyper parameter of the precision matrix Q. Assuming independence from other variables of the model, the optical flow is thus given as the posterior

$$p(\mathbf{u}, \mathbf{v}|\nabla\eta) \propto \int p(\nabla\eta|\mathbf{u}, \mathbf{v}) p(\mathbf{u}, \mathbf{v}|\alpha) p(\alpha)\, \mathrm{d}\alpha$$

for which inference including hyper parameter integration is straight forward using the INLA technique described in Section 2.7.

4.2.2 Experiments & Results

This experiment is dedicated to a first attempt of flow estimation and a subsequent detection of dust sources. It is based on the dust event and respective LPF predictor presented in Section 3.2.4. The last frame of the sequence we are computing the Bayesian optical flow for is shown in Figure 3.6(a). In Figure 4.2(a) the maximum a posteriori optical flow for this frame and its predecessor are shown. In prospect of determining the source of this dust plume we determine the optical flow solely on the positive part $\hat{\eta}$ of the predictor. That is,

$$\hat{\eta}(x, y, t) = \eta(x, y, t) \cdot \mathbb{I}_{\eta>0},$$

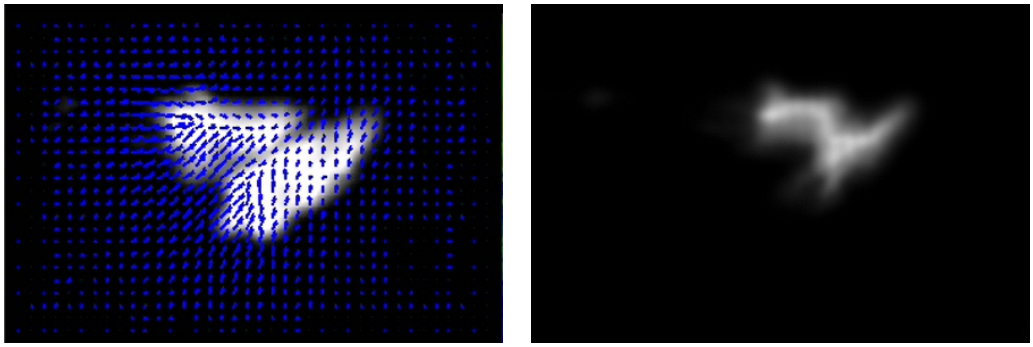
where $\mathbb{I}$ is the indicator function. This reduces the influence of dust-free regions on the optical flow and puts an emphasis on likely dusty areas. On the resulting spatio-temporal sequence of linear predictors starting at $t = 1$ and ending at $t = T$ we perform the following procedure. Let $\hat{\eta}_t = \hat{\eta}(x, y, t)$ with $(x, y) \in \Omega$ denote a frame of the predictor sequence at time t and let $\beta_T = \hat{\eta_T}$ as well as $t = T$.

1. Compute the maximum a posteriori estimate $\mathbf{w}_t = (u, v)_t$ of the optical flow for $(\hat{\eta}_t, \hat{\eta}_{t-1})$.
2. Set $\beta(x, y, t-1) := \beta(x+u, y+v, t)$ in accordance with the brightness constancy equation 4.17 and using bilinear interpolation for non-integer coordinates.
3. If $t = 2$ end the procedure, go back to step 1. otherwise.

The result $\beta(x, y, 1)$ of this approach is shown in Figure 4.2(b). The procedure of using the BCE (or similar constraints) to transform a given image is commonly referred to as *warping*. Now, if the predictor is interpreted similar to a tracer injected into the atmosphere[1], $\beta(x, y, 1)$ can be interpreted as a relative assertion about the strength of the local dust sources.

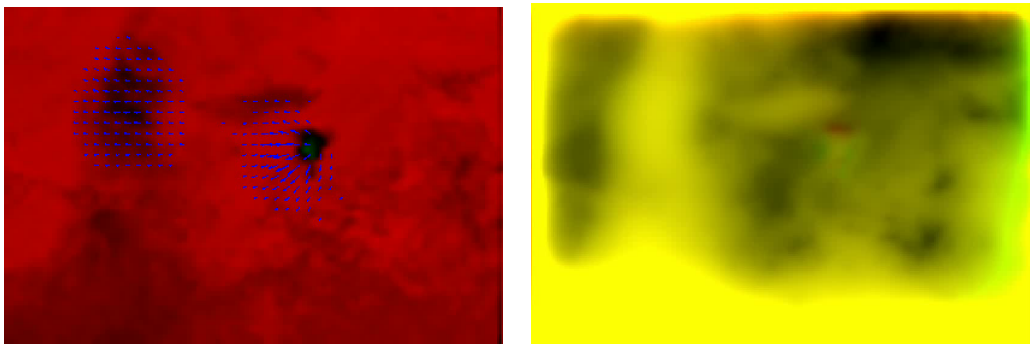
One of the benefits of using the INLA technique for inference is the immediate availability of posterior flow field marginals. These are shown in Figure 4.3 for an early stage of the dust plume considered. Clearly, areas close to the boundary and those with a predictor values close to zero exhibit increased uncertainty in both flow field components (depicted by yellow, a homogeneous mixture of green and red). Regions with inhomogeneous gradient information, e.g. in the center of the image showing the emerging plume, can be identified by red or green coloring. Here, the aperture problem has to be kept in mind. Also, not the spatial proximity of the emerging plume to our estimated source region. On the right hand side of the plotted marginals a green area is visible. This can be contributed to a big but extremely faint plume only slightly noticeable when visually examining the dynamics of the image sequence.

[1] This is a common approach in field experiments to determine areas of dust deposition



(a) (b)

Figure 4.2: First results on incompressible flow estimation and source detection. Panel (a) shows the linear predictor of a dust plume analyzed in Section 3.2.4 together with its (incompressible) flow field (blue arrows). Warping back the predictor according to the corresponding sequence of flow fields results in the source map shown in Panel (b).



(a) (b)

Figure 4.3: Marginal posterioirs of an incompressible flow field. Panel (a) shows the LPF predictor of the dust plume analyzed in Section 3.2.4 at an early stage. The predictor is depicted as red and green where it is smaller where it is lower and greater than zero, respectively. In Panel (b) the brightness of the red and green channel are proportional to the variances of the horizontal and vertical flow field marginals, respectively.

4.2.3 Conclusion

We describe a Bayesian hierarchical model that embeds the well known method of Horn & Schunck into a probabilistic context. Thereby, the tuning parameter of the method receives a natural interpretation as a hyper parameter of a precision matrix parameterizing a latent GMRF that represents the optical flow. Inference withing this framework is straight forward using the INLA technique and we showcase its benefits. We are able to determine the optical flow induced by a larger dust plume and rewind its motion such as to point out the region it was emitted from.

However, considerable work remains. The linear predictor of the LPF method does not necessarily reflect a physical quantity like the mass of dust or optical thickness in a vertical column. Corresponding relations are still to be evaluated. Moreover, the assumed constancy of brightness is clearly violated in case of dust storm. The extend of the resulting error in the flow remains to be clarified and alternative approaches to be sought. Lastly, in prospect of a generally sound model and a possible physical interpretation, the process of thresholding the predictor at some arbitrary value is rather dissatisfying. For instance, minimizing the effect of background noise from regions not affected by dust could supersede this necessity.

4.3 Linear Versus LSM Predictors for Optical Flow

In Section 3.3, which presents results published by Bachl et al. (2013a), the LSM method for dust detection was introduced. A main motivation to develop this method was that approaches based on linearly projecting SEVIRI data inevitably propagate the signal intrinsic noise to the predictor. In particular, dust predictors of pristine regions exhibits a level of variance that hamper the estimation of optical flow. Thresholding the predictor, as proposed by Bachl and Garbe (2012) in line of tracking back dust plumes to their source, however, restricts flow estimation to dense dust plumes. Thin or early dust activity for which the variance itself is an indicator might thereby be completely excluded from the analysis.

In this contribution the effect of employing the LSM method and consecutively estimating the optical flow of a dust plume is compared to motion estimation on top of linear approaches. The presented results were published together with the variance analysis presented in aforementioned Section 3.3. In Section 4.3.1 we provide a concise summary of the methods in focus. The experiment and respective results are elaborated on in Section 4.3.2 and concluded upon in Section 4.3.3.

4.3.1 Methods

Three dust predictors are employed for the upcoming study: LDA, LPF and LSM based on background estimates. For details see Section 3.1.3, 3.2 and 3.3, respectively. Based upon these we estimate the optical flow employing exactly the BHM constructed in Section 4.2. In addition, for reasons of comparison, we determine the optical flow directly from the SFI. This is done by leveraging the optical flow BHM to combine three likelihoods per spatial location. (x, y). That is,

$$
\begin{aligned}
u\,\frac{\partial R}{\partial x} + v\,\frac{\partial R}{\partial y} &= -\frac{\partial R}{\partial t} + \epsilon,\\
u\,\frac{\partial G}{\partial x} + v\,\frac{\partial G}{\partial y} &= -\frac{\partial G}{\partial t} + \epsilon,\\
u\,\frac{\partial B}{\partial x} + v\,\frac{\partial B}{\partial y} &= -\frac{\partial B}{\partial t} + \epsilon,
\end{aligned}
$$

with identical error distributions $\epsilon \sim \mathcal{N}(0, \sigma^2)$ and $\mathbf{I} = (R, G, B)$ being the three channels of the SFI data representation. The posterior of the optical flow is then given by

$$
p(\mathbf{u}, \mathbf{v}|\mathbf{I}) \propto \int p(\nabla R|\mathbf{u}, \mathbf{v}) p(\nabla G|\mathbf{u}, \mathbf{v}) p(\nabla B|\mathbf{u}, \mathbf{v}) p(\mathbf{u}, \mathbf{v}|\alpha) p(\alpha)\, \mathrm{d}\alpha,
$$

for which again inference using INLA is straight forward.

4.3.2 Experiments & Results

We inspect the same two dust events described in Section 3.3.2, which are rather faint and, in the second case, appear as noisy outside the dust covered regions. Event A, shown in Figure 4.4(a), is a five hour sequence of multiple faint plumes with a very smooth trajectory. Event B, shown in Figure 4.4(b), is composed of two hours of multiple small plumes coming from comparably close but distinguishable source areas.

In order to get an idea of how the described noise enters the optical flow we compare the standard deviation of the flow magnitude along the event sequences for flow based on LDA, LPF and LSM predictors as well as the three channels of the SFI representation. As can be seen from Figure 4.4, a projection according to LDA (Panels (e) and (f)) is not necessarily improving the optical flow with respect to dust activity monitoring compared to flow directly computed from SFI data (Panels (c) and (d)). Though LDA improves the signal of the faint plume in the upper region of event A, seemingly dust free regions at the outer regions of both inspected regions reveal flow induced by the background. Results for the LPF method are omitted since

they suffer from similar problems, even though the aforementioned plume induces a slightly better signal than in the case of LDA. Contrarily, the magnitude variance of the flow determined from the LSM predictor (Panels (g) and (h)) does not seem to suffer from this imbalance. Dusty regions clearly show flow activity while dust free regions exhibit only slight variations.

4.3.3 Conclusion

Our results show that the proposed LSM provides a predictor that yields improved estimates of optical flow. In particular, the propagation of the signal intrinsic noise is greatly reduced in regions without dust activity and thereby mitigates the influence these regions have on the optical flow. Dusty regions, on the other hand, are amplified by the LSM method. Here, dust motion is captured that would otherwise have been missed.

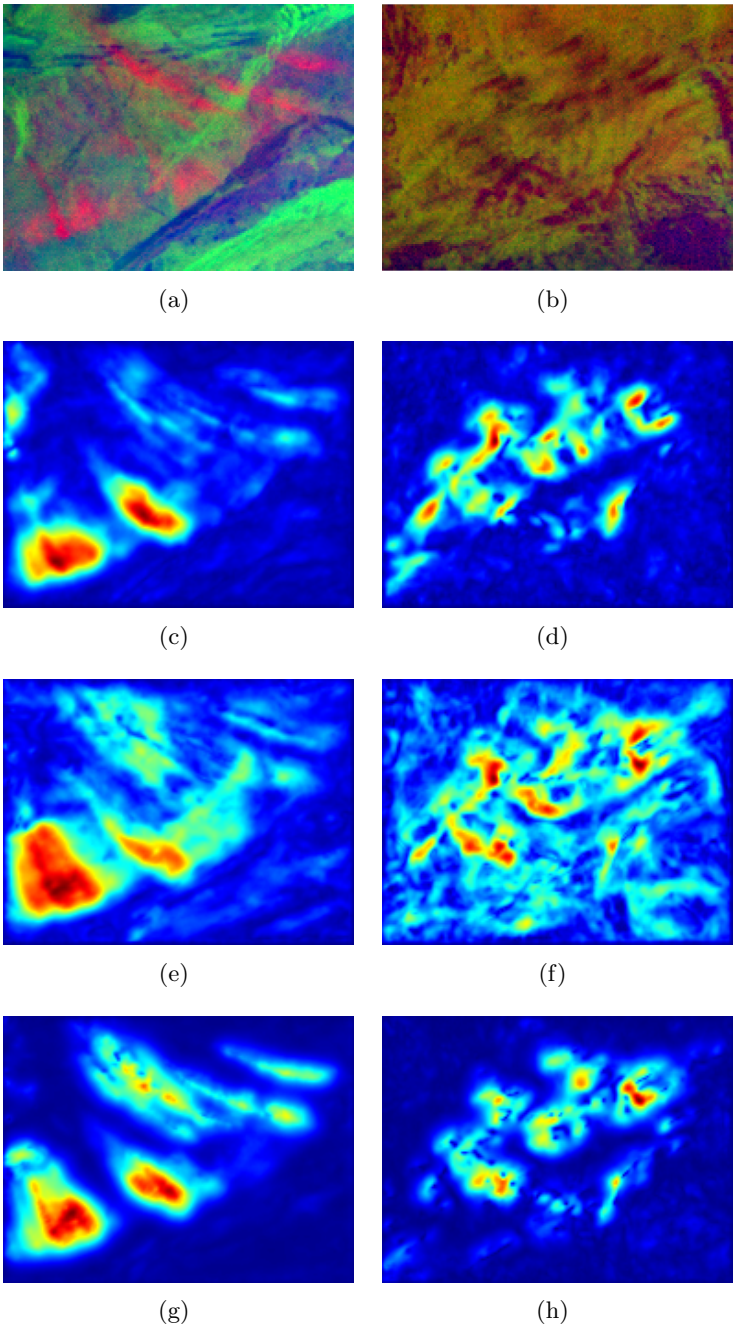


Figure 4.4: Influence of detection methods on optical flow estimates. Spatially normalized temporal standard deviation of the flow magnitude of two dust events are shown in the left and right column, respectively. Panels (c) and (d) refer to the optical flow determined jointly from the channels of the SFI. Panels (e) and (f) as well as (g) and (h) refer to the flow based on LDA and LSM, respectively.

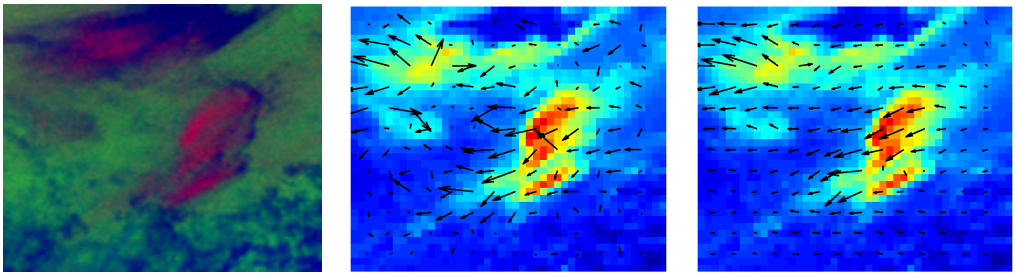
4.4 A BHM for Compressible Optical Flow

While Horn and Schunck optical flow is sufficient to model motion of rigid bodies in many areas of image processing, it is insufficient in capturing the dynamics of aerosols. The assumed constancy of image brightness implies that the flux of the quantity under consideration is divergence-free. This can be seen when comparing the data term requirement $\eta_t + u\eta_x + v\eta_y = 0$ with the continuity Equation (4.4) and its simplified counterpart for incompressible fluids in Equation (4.7) (see Section 4.1.1). The atmosphere and thus dust aerosol, however, is a gaseous mixture that is known to exhibit compressible flow dynamics. Moreover, imaging techniques like SEVIRI deliver a two-dimensional representation of a three-dimensional physical process, which leads to two problems. Firstly, this is not represented in the derivation of the HS approach and the physical interpretation is unclear. Secondly, even if the quantity for which we yield to determine the motion obeys exclusively incompressible flows in the three dimensional space, the reduction to two dimensions may lead to effects that resemble compressible behavior.

In the following it is elucidated how to set the previously developed BHM (Section 2.4) for incompressible optical flow (Section 4.2) forth to perform successful inference on compressible flows. Figure 4.5, published by Bachl et al. (2013a), shows an early result for this Bayesian integrated continuity equation (ICE) method. In 4.5(c) it depicts the posterior mean of the ICE method applied to the outcome of our LSM detection method with respect to the scenery shown in Figure 4.5(a). A comparison to the results of the HS method in Figure 4.5(b) suggests that the former is, at least visually, more appropriate as it is considerably more consistent and suffers from less outliers close to the sources of the dust plumes. This does not come as a surprise since in particular the effect of material entering the imagery is in clear dissent to the BCE formulation.

The ICE method itself is derived in Section 4.4.1. Estimation of the posterior distributions for the flow vector fields under the ICE paradigm is again easily performed using the INLA methodology elaborated on in Section 2.7.

We then proceed with a series of studies that were partly published by Bachl et al. (2013b) and in whole by Bachl et al. (2013a). Section 4.4.2 conducts a simulation study (since ground truth of vector fields is unavailable) that assesses the performance of the ICE formulation of optical flow over the original HS formulation. In Section 4.4.3 an in-depth investigation of two dust storms is presented and we show how our method is able to correctly identify the storm, model its flow and infer aspects of its source. We then demonstrate the forecasting capabilities of the ICE formulation on the basis of a large scale dust event featured in Section 4.4.4. As Section 4.4.5 shows, these forecasts can be improved upon by means of the Bayesian approach we take, i.e. by postprocessing using marginal flow densities. Finally, we conclude in



(a) (b) (c)

Figure 4.5: A first comparison of compressible and incompressible flow. Panel (a) shows an emerging dust plume close to the Bodélé depression and moving to the West as well as slightly to the South ($SFI_{\gamma=1}$ visualization). In Panel (b) the corresponding flow field based on the Bayesian HS method is shown. Bayesian inference using the continuity equation for compressible flows results in the field depicted in Panel (c).

Section 4.4.6 with a procedure capable of tracing a dust storm back to its source and indicating the respective emission strength.

4.4.1 Methods

Our approach follows a line of physically motivated arguments brought up by Corpetti et al. (2002) (for details see also references therein) in the context of non-probabilistic water vapor motion estimation. First off, these authors note that in their context the appropriate model for the vapor density ρ is given by the continuity equation

$$0 = \frac{\partial \rho}{\partial t} + \mathrm{div}(\rho \mathbf{V}), \tag{4.29}$$

where $\mathbf{V}$ is the three-dimensional velocity of ρ. At the same time simple models of atmospheric mechanism and thermodynamics can be employed to connect ρ to brightness temperatures and the underlying radiance measurements provided by instruments like SEVIRI. Under the simplifying assumption of a homogeneous temperature profile of the troposphere the brightness can now be linearly converted to the pressure P. The troposphere is the lowest layer of the atmosphere and carries most of the water vapor and aerosols. If additionally a atmospheric static equilibrium is assumed the vertical component of pressure forces and the gaseous weight cancel out. The vertical pressure change is then proportional to the material density ρ, i.e.

$$\frac{\partial P}{\partial z} = -\bar{g}\rho, \tag{4.30}$$

where z denotes the altitude, $\bar{g}$ the gravity (assumed constant). Reconsidering the measured brightness temperature BT one obtains proportionality to the column

integral of the mass density:

$$BT(x, y, t) \propto \int_{z_0}^{z_1} \rho(x, y, z, t)\, \mathrm{d}z. \tag{4.31}$$

Now, as noted by Corpetti et al. (2002), Fitzpatrick (1988) show that the integration of Equation (4.29) over the column coordinate z while plugging in Equation (4.31) leads to

$$\frac{\partial BT}{\partial t} + \mathrm{div}(BT\mathbf{w}) + [BT\mathbf{n}_z \cdot \mathbf{V}]_{z_0}^{z_1} = 0. \tag{4.32}$$

Here, $\mathbf{w}$ is the *apparent* bidimensional velocity field

$$\mathbf{w} = \frac{\int_{z_0}^{z_1} \rho \bar{\mathbf{V}}}{\int_{z_0}^{z_1} \rho}, \tag{4.33}$$

$\bar{\mathbf{V}}$ denotes the first two components of $\mathbf{V}$ and $\mathbf{n}_z^T = [0, 0, 1]$. Lastly, if the vertical component $\mathbf{n}_z \cdot \mathbf{V}$ of the velocity field is neglected (a common assumption in meteorology[2]), Equation (4.32) reduces to a continuity equation with respect to the brightness temperature and the apparent flow velocity.

Throughout the following we will assume that the predictor $\eta(x, y, t)$ provided by the LSM method behaves similar to the brightness temperature and hence the integral of the dust density over the measurement column. Substituting BT with η and performing the aforementioned reduction then gives

$$0 = \frac{\mathrm{d}}{\mathrm{d}t}\eta \quad = \quad \eta_t + \mathbf{w} \cdot \nabla\eta + \eta\, \mathrm{div}(\mathbf{w}).$$

To summarize, the core component of our approach is a continuity equation for dust predictor that implicitly assumes a vertical integration, hence the name integrated continuity equation (ICE) method. Equation 4.34 also shows the connection to the HS method as it reduces to the former for incompressible materials when the divergence of $\mathbf{w}$ fulfills $\mathrm{div}(\mathbf{w}) = 0$.

The effect on the image intensity can also be understood by considering that the ICE implies

$$\eta(x + u, y + v, t + \Delta t) = \eta(x, y, t) \exp(-\mathrm{div}(\mathbf{w})) \tag{4.34}$$

in a discrete setting where Δt is the time between two images and the flow field v is assumed to be constant. Hence if the divergence is zero the image intensity is conserved along the motion trajectory while it is increased or decreased with progressing time for negative and positive values of the divergence, respectively. For details a derivation of this results see Corpetti et al. (2002)[3].

[2] This assumption is obviously not met for dust plumes gaining elevation and will be discussed in Chapter 5

[3] Note also that (as opposed to us) these authors employ Equation (4.34) for inference and refer to it as the integrated continuity equation

Given a flow field $\mathbf{w}$ Equation (4.34) is also easily employed to infer the temporal predecessor or successor of a given spatial predictor. In practice this is achieved by bilinear interpolation of the intensity values at non-integer coordinates and a successive scaling according to an approximation of the divergence (see below).

Finally, we define the optical flow w according to the ICE in two ways. In prospect of further analytical considerations in Section 4.5 we say that, for given smoothness parameter α, w is the minimizer of the functional

$$\mathcal{J}_{\mathrm{ICE}}(\mathbf{w}) = \int_{\Omega} (\eta_t + \mathbf{w} \cdot \nabla\eta + \eta \operatorname{div}(\mathbf{w}))^2 + \alpha^2(|\nabla u|^2 + |\nabla v|^2). \tag{4.35}$$

In practice, e.g. the following experiments, we define and determine the flow v as the maximum a posteriori estimate of the following BHM. As with the BHM for HS (incompressible) optical flow we assume CAR priors for the vector components of $\mathbf{w}(x, y, t) = [u(x, y, t), v(x, y, t]$ and a log Gamma prior for the shared smoothness parameter α (see Section 2.5 for details on the Gamma distribution). Using the discrete divergence approximation

$$\operatorname{div}([u, v]_{ij}) \approx \frac{1}{2}\big((u_{i,j+1} - u_{i,j-1}) + (v_{i+1,j} - v_{i-1,j})\big)$$

then leads to the following likelihood equation of the flow field given the image

$$u_{ij}\eta_x + v_{ij}\eta_y + \frac{\eta}{2}\big((u_{i,j+1} - u_{i,j-1}) + (v_{i+1,j} - v_{i-1,j})\big) = -\eta_t + \epsilon_{ij},$$

where again $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$. It should be noted that under both the HS and the ICE method, the scale of motion that can be recovered is limited by the range over which the partial derivatives are computed. A well known remedy is to determine the flow on a pyramid of different scales. For the sake of simplicity we refrain to follow this strategy for the study at hand and determine image derivatives on a resolution of 5 pixels as we found that this suffices to capture large scale flow fields of fast dust storms.

In what follows we show that the Bayesian ICE approach to determining optical flow of dust storms considerably improves estimated flow fields obtained using HS methods, largely for the obvious reasons that dust storms grow and then diminish through time. As should be clear from the development, estimation of the posterior distribution $p(\mathbf{u}, \mathbf{v}|\nabla\eta)$ for the flow vector fields under either the HS or ICE paradigms is easily performed using the INLA methodology (Rue et al., 2009).

4.4.2 A Simulation Study

We now compare the HS method to the ICE method in reconstructing a flow field, both under classical and the proposed Bayesian perspective. Since ground truth is unavailable for the Saharan dust storms, we use a synthetic image sequence to

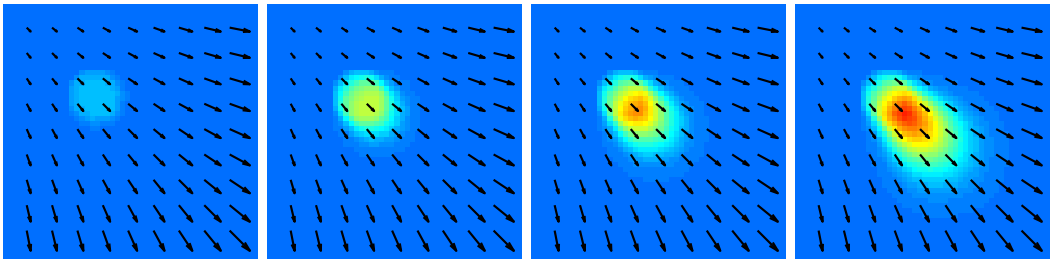


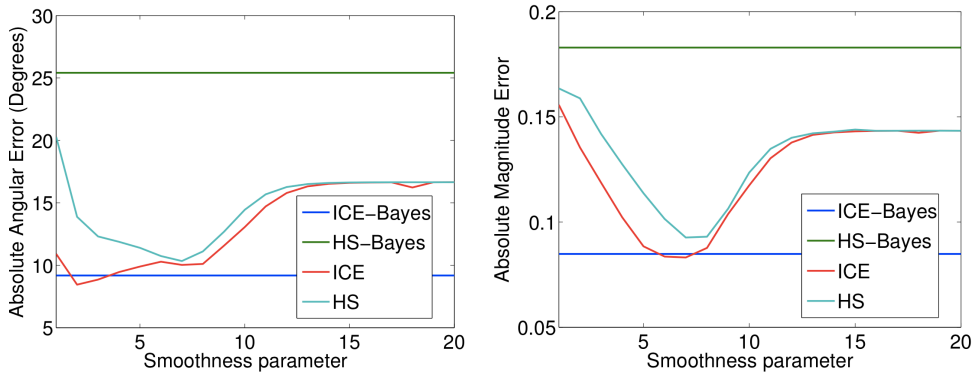
Figure 4.6: Synthetic image sequence of a dust plume and aerosol flow.

illustrate the difference between the two approaches. Figure 4.6 shows the progression we consider, a constant flow field with a growing dust plume.

We assume the location of the dust plume is known and estimate the flow field under HS and ICE based on this sequence. Figure 4.7 shows the mean absolute error in angular (left panel) and magnitude (right panel) estimates for four approaches: ICE and HS where the precision parameter α is set by hand (equivalent to the current best practices) and the corresponding Bayesian approaches where the INLA methodology is used to estimate this parameter. Figure 4.7 shows several interesting features. The first is that for any level of α and any error metric, the ICE approach outperforms the HS approach. This indicates the benefit of using ICE over the BCE when the preservation of brightness assumption is clearly violated.

The second conclusion speaks to the benefit of estimating α via Bayesian methods. In this context we see that, depending on the metric, different choices of α are optimal in case of ICE. However, by intrinsic parameter integration, the ICE method under Bayesian estimation outperforms the regular ICE approach for almost all levels of α, and even at its best the standard ICE method is barely better than the Bayesian approach. Finally, there is an interesting warning regarding model misspecification. We see that the HS method, when estimated by Bayesian methods, performs considerably worse than all other approaches, due to the violation of BCE in our example.

A clear motivation to apply Bayesian hierarchical modeling and inference with INLA is the straight-forward assessment of the marginal distributions of the latent fields. Figure 4.8 shows why this is of great importance, in particular in applied contexts. It visually compares the marginal flow component variances of the Bayesian ICE and HS approach. From Figures 4.8(a) to (c) it becomes obvious that the uncertainty in the ICE flow estimates exactly corresponds to those that are inherent to the model. The outer boundary of the simulated dust source region is colored in either red or green, displaying a high variance in either one of the field components. This is as expected for two reasons. Information about the direction of the motion is obtained from the gradient of the image sequence, which is most prominent in this region and hence leads to low variance in one component. However, due to the aperture



(a) Angular error (b) Magnitude error

Figure 4.7: Angular (a) and magnitudial (b) error of aerosol flow estimation for the synthetic image sequence in Figure 4.6. The plots compare the errors of the ICE and the HS methods under both standard and Bayesian inference as a function of the smoothness parameter α.

problem, motion perpendicular to the gradient can not be detected by the model. As the simulated dust moves, from a local perspective, mostly downwards on the upper and lower boundary of the source and to the right at the right and left boundaries, this leads to the high variance in the horizontal and vertical components, respectively. The outer region as well as inside the dust source behave as expected as well. Inside both no gradient is present, leading to uncertainty in both components and thus in a yellow coloring, which results from mixing red and green. This phenomenon is most present at the borders of the image region. Here, the least information is propagated through the latent GMRF coupling from the central region.

As one can see from Figures 4.8(d) to (f), the marginal variances can be very informative in terms of misspecifications of the model. In 4.8(d) and (e) the Bayesian HS approach contributes the uncertainty either fully to the horizontal or vertical flow field component, a highly undesirable behavior and presumably an effect of the violated brightness preservation assumption. In Figure 4.8(f) the situation is more balanced but still predominance of the uncertainty in the horizontal (red) component can be observed. Lastly, the central area of the dust source region in this figure seems slightly more pronounced than in Figure 4.8(c). This is an additional indicator for the fact that in particular the HS approach struggles to reflect dust source effect, e.g. the influx of dust mass into the atmosphere.

4.4.3 Case Studies

After establishing the good performance of our dust detection routine and the Bayesian ICE method of reconstructing the flow field, we highlight the use of our

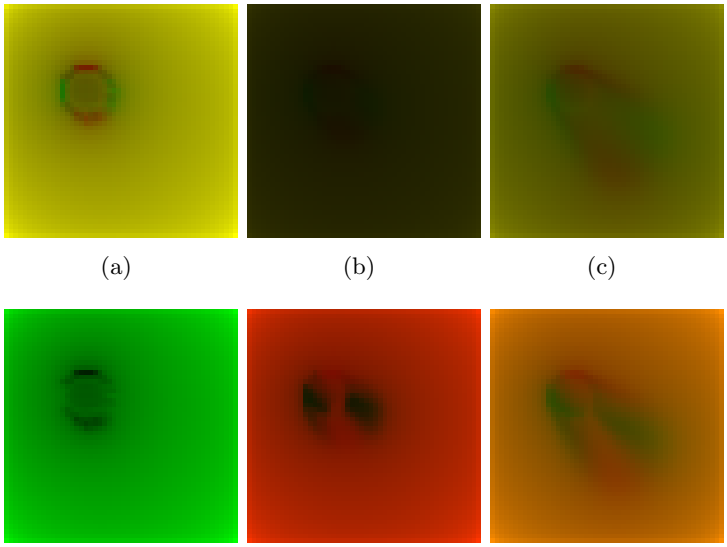


Figure 4.8: Exemplary marginal variances of the flow field vectors derived via the Bayesian ICE ((a) to (c)) and HS ((d) to (f)) approaches. The brightness of the red and green channel are proportional to the variances of the horizontal and vertical field components, respectively.

framework during the evolution of two separate dust storms. Figures 4.9 and 4.10 show dust storms that occurred during January 8, 2010 and January 16, 2010, respectively. The figures show the pixel-wise probability of dust estimates under the emissivity LSM approach and, furthermore, compare the estimated flow fields under the Bayesian HS and the Bayesian ICE approaches.

We see several features from Figures 4.9 and 4.10. The first is that the detection appears to be working well. Points which are clearly dusty are correctly given high probabilities, while the model captures uncertainty in the estimates around the edges of the dust plumes. Secondly, we see why the ICE method is preferred over standard HS. There is considerably more regularity to the estimated flow field in the third row than the second row, especially in the first two time points. This enables a coherent reconstruction of the dust plume flow. Furthermore, the Bayesian HS method seems unable to detect the flow of smaller dust storms, such as the one featured in the lower right hand corner of the plots in Figure 4.9.

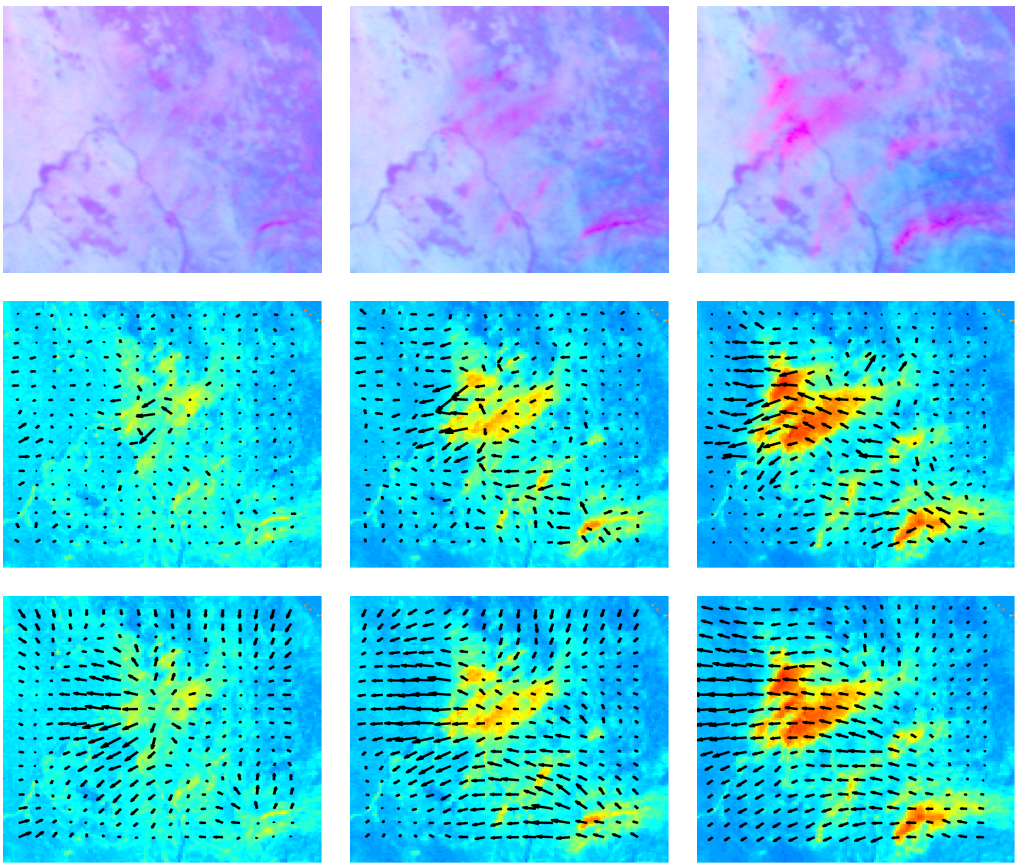


Figure 4.9: Dust plume on January 8, 2010 at 7.15 am, 8.30 am and 11 am GMT. Top row: observed satellite data in false color; middle row: pixel-wise LSM probability of dust estimates (with high probabilities in red) overlaid with the Bayesian HS flow field; bottom row: same pixel-wise probability of dust estimates as above now overlaid with the Bayesian ICE flow field.

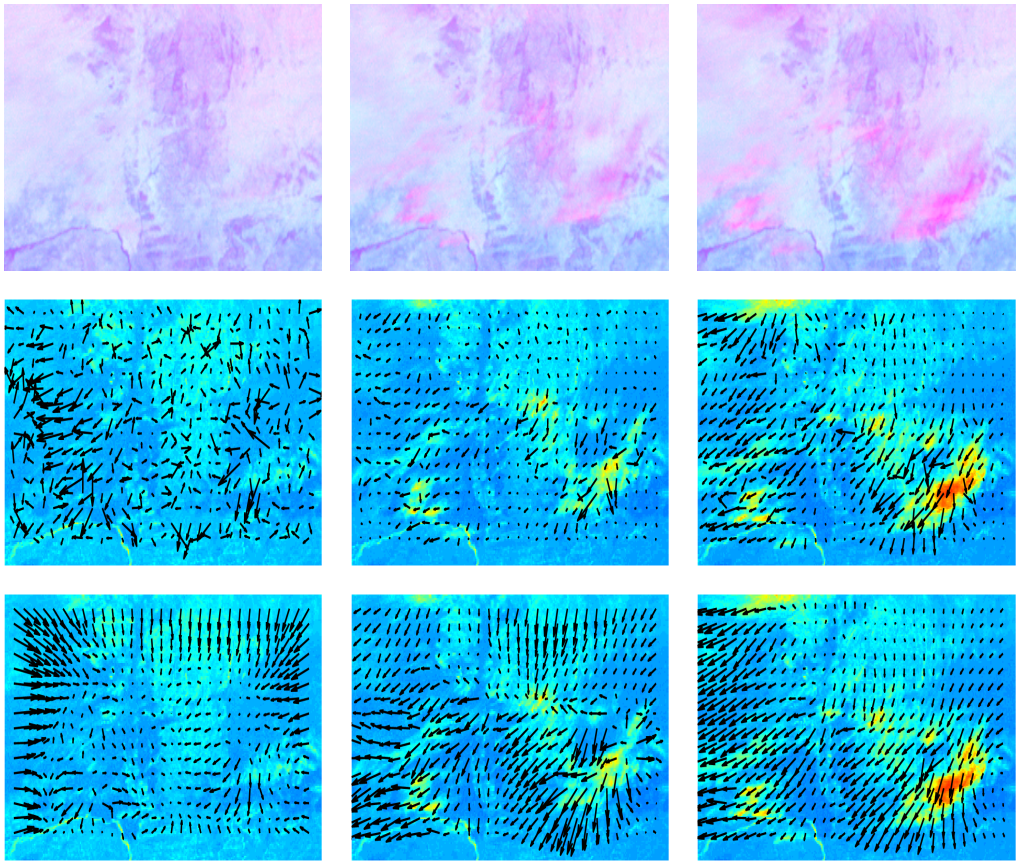


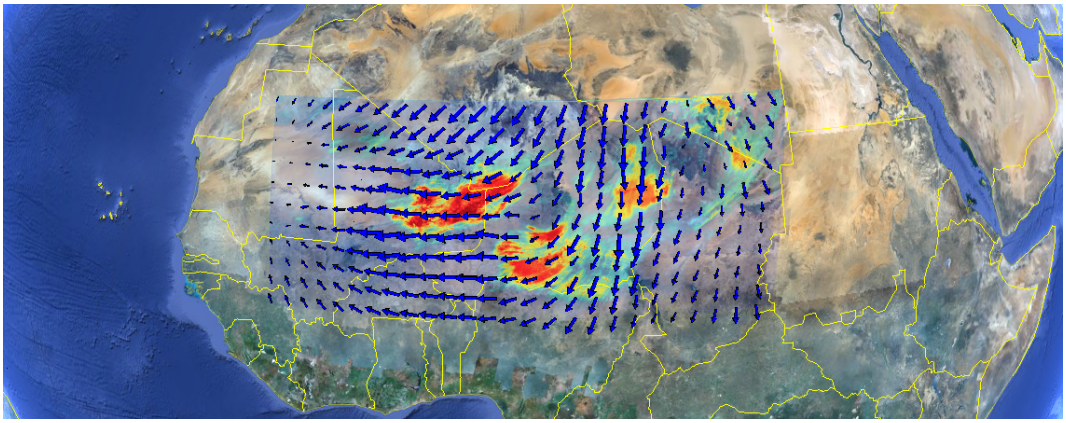
Figure 4.10: Dust plume on January 16, 2010 at 10.15 am, 11.45 am and 1 pm GMT. Top row: observed satellite data in false color; middle row: pixel-wise LSM probability of dust estimates overlaid with the Bayesian HS flow field; bottom row: same pixel-wise probability of dust estimates as above now overlaid with the Bayesian ICE flow field.

4.4.4 Forecasting Dust Events

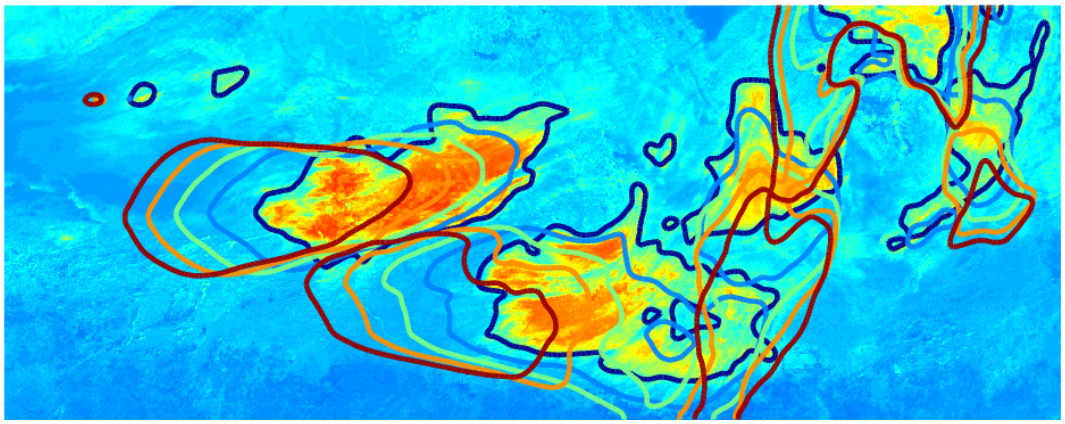
We now discuss an application of our method that is relevant for areas in proximity of regions that emit dust. In a study reflected by Figure 4.11 we focus on the area surrounding a massive dust storm occurring on January 17, 2010 and a respective assessment of the risk to be affected by it. Equation (4.34) implies a straight-forward method of extrapolating the future development of a spatial dust density estimate given a flow field one time step ahead. This can be employed in an iterative scheme.

Firstly, we compute the dust predictor and flow with respect to the imagery of 11:45h and 12:00h GMT. Figure 4.11 (a) shows the outcome of this procedure as an overlay to the earth surface imagery as shown by Google Earth. Three dust plumes of large size are clearly visible: (A) One over north eastern Niger predominantly moving to the south, (B) one over southern Niger moving in southern and western direction and lastly (C) a plume emerging at the borders between Algeria, Niger and Mali moving westwards in direction of central Mali. Then we extrapolate the field through iterative application of Equation (4.34) for 96 steps, under the assumption that the given flow field remains approximately constant throughout this time period. As one time step in the SEVIRI imagery corresponds to 15 minutes of time difference, this results in an estimate of the dust density development 24 hours ahead. One can now make use of this forecast to predict the future location of the main body of dust mass. Figure 4.11 (b) shows the 30% dust density contour lines of the initial imagery as well as for forecasts of 6, 12, 18 and 24 hours ahead. It is easy to see that these forecasted contours develop according to the estimated flow field. While plume (A) predominantly moves towards the south, (B) and (C) mostly move to the west. Our method thus results in an informative large scale directional assessment of the dust plume development.

Figure 4.12 shows how our forecast gains accuracy with a decreasing prediction interval. The actual dust density at January 18, 2010 12.00h GMT is shown in Figure 4.12 (a). During the morning of this day, three new dust events emerge that mix with the large plume from the day before. These are not yet reflected in the 24h forecast depicted in Figure 4.12 (b), from which another observation can be made. The forecast of Plumes (B) and (C) does not appear as wide spread as the actual outcome. The reason for this is underestimation of the flow magnitude in the western part of the imagery. This causes the forward projection of the dust density to slow down and accumulate mass in this region. Plume (A) shows a similar effect as can be seen from visual inspection of the falsecolor data (not shown here), where it actually leaves the region depicted for this study at the southern border while undergoing a large scale spreading effect that indicates a corresponding wind field during the night. However, the 3h and 1h forecasts shown in Figures 4.12 (c) and 4.12 (d) gain accuracy. Both indicate the new dust event in the south east and the 1h forecast also picks up the two weaker events.

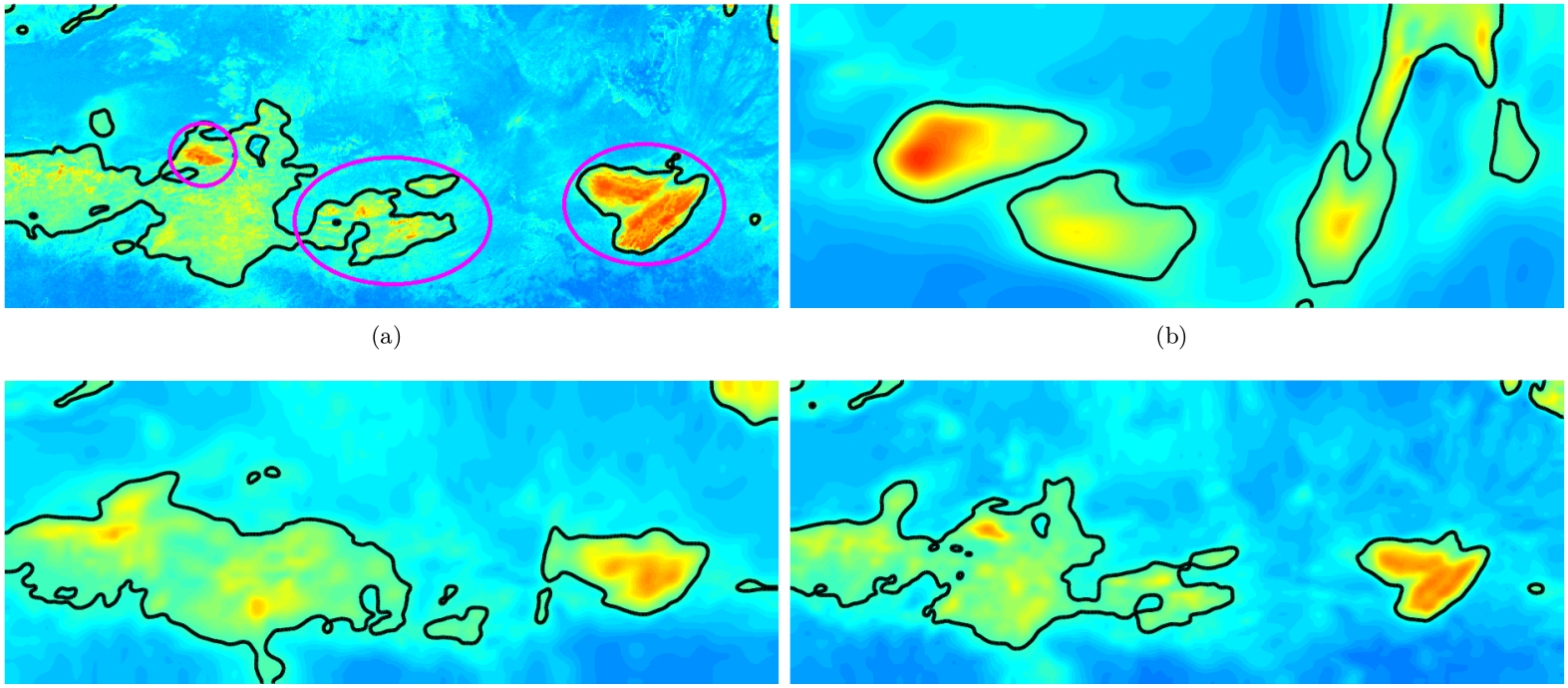


(a)



(b)

Figure 4.11: Forecast of a dust event. Panel (a) shows the linear predictor of a dust event over northern Africa on January 17, 2010, 12h GMT as a transparent overlay on Google Earth imagery including country borders and the flow field computed with the ICE method. Panel (b) depicts the contour lines of 30% dust density according to the forecast derived from the data in (a) for 0 (dark blue, the data itself), 6 (light blue), 12 (green), 18 (orange) and 24 (red) hours after the event. Note that the detection in the top right corner is a false positive due to influence of a water cloud.



(a) (b)

(c) (d)

Figure 4.12: Forecasted dust densities. Panel (a) shows the actual predictor for dust density on January 18, 12h GMT with a 30% contour line (black). The pink ellipses mark dust plumes emerging during the morning of this day. Panels (b), (c) and (d) show the dust density forecasted for this time based on data from 24, 3 and 1 hour before, respectively.

4.4.5 Marginal Posteriors for Forecast Postprocessing

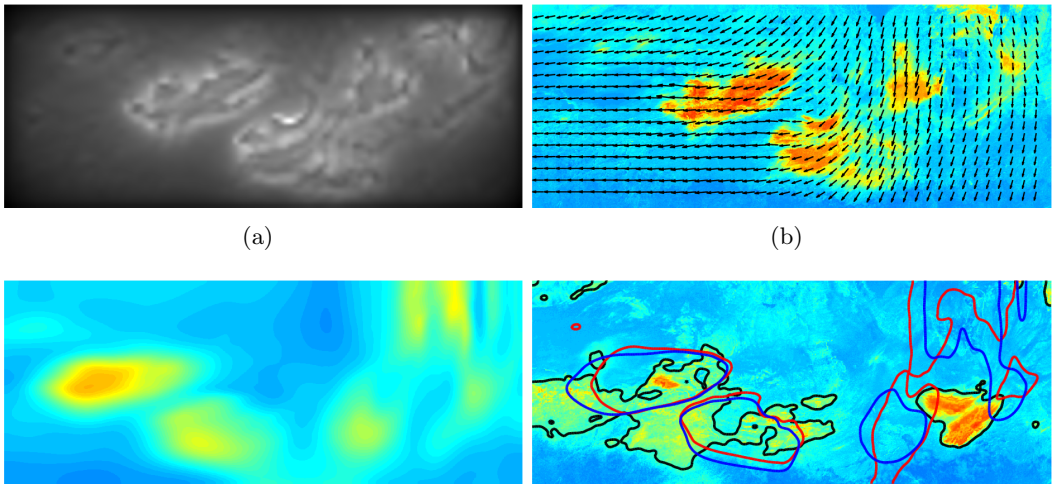
As shown in the previous section, missing or noisy predictor gradient information can lead to poor forecasts. However, the acquired marginal posteriors of the flow field offer valuable information on where this is the case. We will now show a simple but effective way to make use of this information to alleviate the effects of non-informative regions. Consider again the dust event of January 17, 2010 and Figure 4.13 (a) that shows the per-pixel sum of the estimated flow component posterior precisions derived with the INLA method. Clearly, most precision is obtained at the borders of the dust plumes where the gradient has a sufficient magnitude. This fact can be used to re-estimate the flow field via a spatially weighted variant of the ICE method:

$$L_{\mathrm{ICE^W}}(\alpha) = \int_\Omega (q(\gamma_1 + \gamma_2))^2 (\eta_t + \mathbf{w} \cdot \nabla\eta + \eta \operatorname{div}(\mathbf{w}))^2 + \alpha^2(|\nabla u|^2 + |\nabla v|^2),$$

where γ_1 and γ_2 are the respective precisions and q is a fixed factor that scales the set of local precision to the range $[0, 1]$. Figure 4.13(b) shows that this procedure has the intended effect. While the main direction and curvature of the flow field in regions with dust activity is approximately the same as for the unweighted ICE estimates, significant regularity of the field outside these regions is obtained. This effect is most dominant in the western and south eastern part of the area under investigation but also in between dust plumes. Figures 4.13 (c) and 4.13 (d) show visibly that this aides the forecast process and mitigates aforementioned problematic effects. In particular, plumes (B) and (C) move faster towards the west and the mass accumulation effect of plume (B) that occurred with the unweighted method is decreased. The increase of the magnitude of the flow field in the south eastern region leads to a similar observation with respect to plume (A). However, when comparing the forecasted densities of the weighted and unweighted method with the true observation, a general underestimation of the dust motion speed is still apparent. This does not come as a complete surprise as with increasing age dust plumes dissipate into higher altitudes. In these heights the wind speed is most often larger than at ground level. It is therefore highly likely that the underestimation of the motion speed is due to the early stage of the plumes compared to forecast horizon.

4.4.6 Source Detection

In the previous sections a flow field served to predict the future development of a given dust plume. Given a sequence of flow fields, the same idea of transporting the dust plume can be applied in the reverse direction. This way, the mass of the plume is moved to the regions it emerged from and can be used as an estimator of the respective local emission strength.



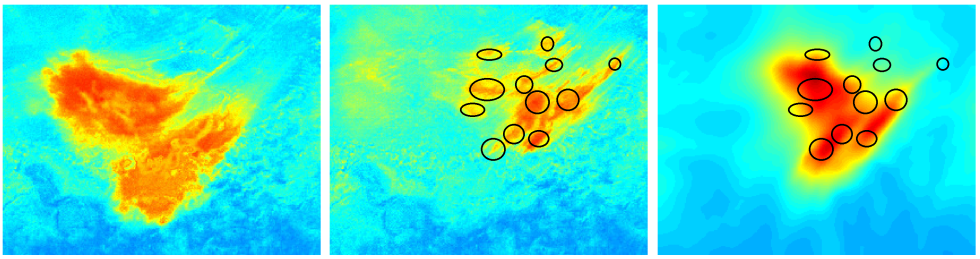
(a) (b)

(c) (d)

Figure 4.13: Postprocessing of flow estimates. Panel (a) shows the sum of the marginal flow component precisions derived for the dust forecast procedure elaborated in Section 4.4.4. These are be employed to weight the local linear fit term of the ICE method, which leads to the flow field shown in Panel (b) and the 24 hour forecast depicted in (c). For comparison (d) shows the actual dust density and 30 % density contour line (black) after 24 hours together with the contour lines of the unweighted (red) and the weighted (blue) forecast method.

Such an estimate is of great interest in environmental sciences. For instance, Jickells et al. (2005) note that the mineralogical composition of a dust plume is inherited from its source region and determines properties such as nutrition effects on terrestrial and marine ecosystems on a global scale. Yet, in-situ measurement sites in Africa are sparsely distributed and data such as horizontal visibility from synoptic stations is hardly sufficient for the identification of source areas (Mahowald et al., 2005). There are, however, studies that employ dust indicators like aerosol depth measured by satellites and perform a long term temporal averaging of this quantity to identify sources. Intrinsically, this leads to overestimating the source strength of regions that are only traversed by dust plumes. This is demonstrated by an experiment of Schepanski et al. (2012) where dust plume trajectories and source regions were determined by human experts visually inspecting SEVIRI imagery. Our method not only yields an automation of this procedure but also compliments other data driven approaches, e.g. studies that rely on wind field averages and Lagrangian trajectories to trace back dust to its origin (Alonso-Pérez et al., 2012).

Figure 4.14 shows the result of the proposed method applied to a massive dust plume occurring on January 18, 2010 over the Bodélé depression in northern Africa. First



(a) (b) (c)

Figure 4.14: Spatial estimation of dust emission strength. Panel (a) shows the linear predictor of a dust even on January 18, 2010 at 15.00h GMT over the Bodélé depression in northern Africa. This dust plume originates from a cluster of source regions identified by visual inspection of the image sequence and marked with black circles in Panel (b). The flow field estimated from this sequence is used to transport back the dust density in (a) to the presumed origin shown in Panel (c) and serves as a spatial estimator for the dust emission strength.

signs of the event are visible in the data at 6.15h GMT and the plume reaches its maximal extent at around 15.00h GMT. We compute the flow of the plume for the whole period and then use Equation (4.34) to transport the predictor of the imagery at 15.00h GMT (see Figure 4.14 (a)) back according to these estimates. In order to judge the accuracy of the estimated source regions an extensive visual inspection of the linear predictor sequence over time was performed. The black circles in Figure 4.14 (b) mark regions that can be recognized as actually emitting dust rather than just being covered by the plume over the course of time. The linear predictor shown in this Figure represents dust activity at 8.15h GMT where the most active source regions are still identifiable as distinct areas. Note that for instance the source at the very south of the active region appears rather faint but can be clearly identified to emit a large amount of dust when inspecting the dynamics of the image sequence.

Transporting the dust backwards according to the determined flow field leads to the emission strength estimate depicted in Figure 4.14 (c), which is again superimposed with the source region markers. Most interestingly, almost all markers lie within the bulk of the area estimated to have a high emission strength. Vice versa, the emission strength is low outside the cluster of these markers. The only sources that are not captured well are those in the north western corner of the imagery. However, as can be seen from the data, these sources are rather weak and have another property that makes their flow estimation challenging. The spatial extent of all three sources is narrow orthogonal to their motion direction. The imagery gradient necessary for our flow estimation technique thus has a small spatial extent as well and is likely to be too weak to pick up the correct genesis of the plume.

4.4.7 Conclusion

We outline a Bayesian framework leveraging the recently developed INLA methodology for detecting and tracking dust storms. The approach makes several developments, including a superior dust detection methodology, a link between the classical literature of optical flow and Gaussian Markov random fields (GMRFs) and the use of the ICE to model flow fields where an assumption of brightness constancy is inconsistent with the physical process.

In a simulation study we show the improved performance of our Bayesian ICE model over existing procedures. The study emphasizes that Bayesian estimation can alleviate issues related to the setting of tuning parameters and how marginal posteriors are an informative factor in terms of coherence between intrinsic model assumptions and the determined flow field.

Real world examples show the implications of the improvement and a multitude of applications are pointed out. A technique for dust storm forecasts is demonstrated and it is shown how to improve these forecasts by including previously estimated marginal posteriors into the model for flow estimation. Lastly, by rewinding a given dust plume in time we are now are now able to successfully pinpoint its source and an indicator for the respective emission strength.

4.5 Variational Properties of Optical Flow

In the previous sections we determined the flow field of a given material density by minimizing functionals that incorporate variants of the continuity equation. This was done by discretizing the functional such that its minimizing argument approximately equals a maximum a posteriori estimate of a BHM. Accordingly, the variational perspective on solving these problems (see Section 4.1.2) allows for assertions that are of use in the probabilistic context as well.

As shown by Schnörr (1991) and explained in Section 4.1.3, existence and uniqueness of the minimizing function can be guaranteed under mild assumptions on given image data. Moreover, as a consequence of the Lemma of Lax-Milgram (Lemma 3 in Section 4.1.2), a suitably chosen discretization basis spanning a finite-dimensional subspace of functions results in a positive definite linear equation system. As a consequence, the corresponding discretized quadratic functional defines a proper GMRF with the unique minimizer as mode, and with the linear system matrix as precision matrix.

In the following, we present three contributions to a more detailed analysis of functionals involving the continuity equation. In Section 4.5.2 a proof is provided that generalizes the result of Schnörr (1991), (Theorem 7, 4.1.3) to compressible

flows[4]. Hence, one obtains the same guarantees in terms of existence and uniqueness of the flow minimizing the respective functional (see Section 4.4.1) as well as well-posedness of its discrete representation. For future research and completeness of the analysis in general, Section 4.5.3 is concerned with the derivation of the Euler-Lagrange equations that arise from the functional defining compressible flow. Lastly, in Section 4.5.4, we derive the Euler-Lagrange equations of the material density when either a compressible or an incompressible flow field are given. This enables us to point out similarities to the class of SPDEs elaborated on in Section 4.1.4.

4.5.1 Notation

Let us reiterate the functional defining the compressible flow field for given material density η:

$$\mathcal{J}_{\mathrm{ICE}}(\alpha) = \int_\Omega (\eta_t + \mathbf{w} \cdot \nabla\eta + \eta \operatorname{div}(\mathbf{w}))^2 + \alpha^2(|\nabla u|^2 + |\nabla v|^2).$$

In the following, $\mathbf{u} = [u_1, u_2]$ and $\mathbf{v} = [v_1, v_2]$ will denote two flow fields and we use integers to index the horizontal and vertical components, respectively. We define a linearform f_{ICE} and a bilinearform a_{ICE} that constitute the functional $\mathcal{J}_{\mathrm{ICE}}$ in a way that allows us to apply the lemma of Lax-Milgram. For convenience we will use the following notation:

$$\begin{aligned}
\mathbf{d} &= [d_1, d_2]^T = [\partial u_1/\partial x, \partial u_2/\partial y]^T, \\
\mathbf{e} &= [e_1, e_2]^T = [\partial v_1/\partial x, \partial v_2/\partial y]^T, \\
\boldsymbol{\eta} &= [\eta, \eta]^T, \\
\boldsymbol{\eta}_t &= [\eta_t, \eta_t]^T,
\end{aligned}$$

The bilinear form corresponding to the Horn and Schunck approach is composed of the components

$$\begin{aligned}
a_{\mathrm{HS}}(\mathbf{u}, \mathbf{v}) &= \int_\Omega (\nabla\eta \cdot \mathbf{u})(\nabla\eta \cdot \mathbf{v}), \\
a_{\mathrm{S}}(\mathbf{u}, \mathbf{v}) &= \int_\Omega \alpha^2(\nabla u_1 \cdot \nabla v_1 + \nabla u_2 \cdot \nabla v_2)\, \mathrm{d}\mathbf{x},
\end{aligned}$$

which correspond to the quadratic parts of the data and smoothness term, respectively. Likewise, the bilinear form for compressible flow can be written as

$$\begin{aligned}
a_{\mathrm{ICE}}(\mathbf{u}, \mathbf{v}) &= \int_\Omega (\boldsymbol{\eta} \cdot \mathbf{d})(\nabla\eta \cdot \mathbf{v}) + (\nabla\eta \cdot \mathbf{u})(\boldsymbol{\eta} \cdot \mathbf{e}) + (\boldsymbol{\eta} \cdot \mathbf{d})(\boldsymbol{\eta} \cdot \mathbf{e})\, \mathrm{d}\mathbf{x} \\
&\quad + a_{\mathrm{HS}}(\mathbf{u}, \mathbf{v}) + a_{\mathrm{S}}(\mathbf{u}, \mathbf{v}).
\end{aligned}$$

[4] Recall that the notion of *(in-)compressible* flow is supposed to reflect whether a method assumes divergence freeness or not. This property is, however, not enforced by any of the approaches.

and the linear form is given by the expression

$$f_{\mathrm{ICE}}(\mathbf{v}) = 2\int_\Omega \eta_t \nabla\eta\cdot\mathbf{v} + \eta\boldsymbol{\eta}_t\cdot\mathbf{e}\,\mathrm{d}\mathbf{x}.$$

In case of $\mathbf{u}=\mathbf{v}$ we also note that

$$\begin{aligned} a_{\mathrm{ICE}}(\mathbf{v},\mathbf{v}) &= \int_\Omega (\nabla\eta\cdot\mathbf{v}+\eta\cdot\mathbf{e})^2\,\mathrm{d}\mathbf{x} + a_{\mathrm{S}}(\mathbf{v},\mathbf{v}) \\ &= \underbrace{\int_\Omega (\nabla\eta\cdot\mathbf{v})^2}_{a_{\mathrm{HS}}(\mathbf{v},\mathbf{v})} + \underbrace{\int_\Omega 2(\eta\cdot\mathbf{e})(\nabla\eta\cdot\mathbf{v})}_{a_{\mathrm{M}}(\mathbf{v},\mathbf{v})} + \underbrace{\int_\Omega (\eta\cdot\mathbf{e})^2}_{a_{\mathrm{Q}}(\mathbf{v},\mathbf{v})}\,\mathrm{d}\mathbf{x} + a_{\mathrm{S}}(\mathbf{v},\mathbf{v}) \end{aligned}$$

where the integral components are

$$\begin{aligned} a_{\mathrm{HS}}(\mathbf{v},\mathbf{v}) &= \int_\Omega \eta_x^2 v_1^2 + 2\eta_x\eta_y v_1 v_2 + \eta_y^2 v_2^2\,\mathrm{d}\mathbf{x}, \\ a_{\mathrm{M}}(\mathbf{v},\mathbf{v}) &= 2\int_\Omega \eta\Big[\eta_x v_{1,x} v_1 + \eta_y v_{1x} v_2 + \eta_x v_{2,y} v_1 + \eta_y v_2 v_{2y}\Big]\,\mathrm{d}\mathbf{x}, \\ a_{\mathrm{Q}}(\mathbf{v},\mathbf{v}) &= \int_\Omega \eta^2\Big[v_{1x}^2 + 2v_{1x}v_{2y} + v_{2y}^2\Big]\,\mathrm{d}\mathbf{x}, \\ a_{\mathrm{S}}(\mathbf{v},\mathbf{v}) &= \int_\Omega \alpha^2(|\nabla v_1|^2 + |\nabla v_2|^2)\,\mathrm{d}\mathbf{x} = \int_\Omega \alpha^2(v_{1x}^2 + v_{1y}^2 + v_{2x}^2 + v_{2y}^2)\,\mathrm{d}\mathbf{x}. \end{aligned}$$

We will also employ the following scalar products and norms encompassing the Sobolev space H and the space of motion fields V. By $(u_1,u_2)_1$ we denote the scalar product

$$(u_1,u_2)_1 = \sum_{|\psi|\le 1} (D^\psi u_1, D^\psi u_2)_0 = \sum_{|\psi|\le 1}\int_\Omega D^\psi u_1 D^\psi u_2\,\mathrm{d}\mathbf{x}$$

and the norm $|u|_1 = [(u,u)_1]^{1/2}$. Here, $(\cdot,\cdot)_0$ denotes the scalar product and norm $|\cdot|_0$ in $L^2(\Omega) = H^0(\Omega)$. The scalar product in V is defined as

$$(\mathbf{u},\mathbf{v})_V = (\mathbf{u},\mathbf{v})_1 = (u_1,v_1)_1 + (u_2,v_2)_1$$

and induces the norm

$$|\mathbf{u}|_V = [(\mathbf{u},\mathbf{u})_V]^{1/2} = \Big[\int_\Omega |\mathbf{u}|^2 + |\nabla u_1|^2 + |\nabla u_2|^2\Big]^{1/2}\,\mathrm{d}\mathbf{x}.$$

Lastly, $|\mathbf{u}|$ denotes the Euclidian length of $\mathbf{u}$,

$$|u|_\infty = |u|_{L^\infty(\Omega)}$$

and $|\mathbf{u}|_0$ means $[(\mathbf{u},\mathbf{u})_0]^{1/2}$, where

$$(\mathbf{u},\mathbf{v})_0 = (u_1,v_1)_0 + (u_2,v_2)_0.$$

4.5.2 Existence and Uniqueness of Compressible Flow

We will generalize the existence and uniqueness results reported in connection with the approach of Horn and Schunck to the more general problem of minimizing L_{ICE}. The major difference concerns the presence of the term $\eta \operatorname{div}(\mathbf{u})$ so as to take better into account compressibility of flows.

Theorem 8 (Existence and Uniqueness)**.** *Let* $H^1(\Omega) = \{u \in L^2(\Omega) : D^\psi u \in L^2(\Omega) \text{ for } |\psi| \leq 1\}$ *denote a Sobolev space and let* $V = \{\mathbf{u} = (u_1, u_2)^T \in H^1(\Omega) \times H^1(\Omega)\}$*. If* $\alpha > 0$*,* $\eta, \eta_x, \eta_y \in L^\infty(\Omega)$*,* η_x *and* η_y *are linearly independent as elements of* $L^2(\Omega)$ *then the functional* $\mathcal{J}_{\mathrm{ICE}} : V \to \mathbb{R}$*,*

$$\mathcal{J}_{\mathrm{ICE}}(\mathbf{u}) = \int_\Omega (\eta_t + \mathbf{u} \cdot \nabla\eta + \eta \operatorname{div}(\mathbf{u}))^2 + \alpha^2(|\nabla u_1|^2 + |\nabla u_2|^2),$$

attains a unique minimum that depends continuously on the image data η*.*

The proof below is based on a series of preparatory Lemmata presented next.

Lemma 4. *The inequality*

$$|a_{\mathrm{M}}(\mathbf{u}, \mathbf{v}) + a_{\mathrm{Q}}(\mathbf{u}, \mathbf{v})| \leq C_{\mathrm{ICE}}(|\mathbf{d}|_0|\mathbf{v}|_0 + |\mathbf{e}|_0|\mathbf{u}|_0 + |\mathbf{d}|_0|\mathbf{e}|_0)$$

holds for some fixed $C_{\mathrm{ICE}} = C_{\mathrm{ICE}}(\eta) > 0$*.*

Proof. Firstly, one has that

$$\begin{aligned} |a_{\mathrm{M}}(\mathbf{u}, \mathbf{v}) + a_{\mathrm{Q}}(\mathbf{u}, \mathbf{v})| &= \Big| \int_\Omega (\boldsymbol{\eta} \cdot \mathbf{d})(\nabla\eta \cdot \mathbf{v}) + (\boldsymbol{\eta} \cdot \mathbf{e})(\nabla\eta \cdot \mathbf{u}) + (\boldsymbol{\eta} \cdot \mathbf{d})(\boldsymbol{\eta} \cdot \mathbf{e})\, \mathrm{d}\mathbf{x} \Big| \\ &\leq |(\boldsymbol{\eta} \cdot \mathbf{d}, \nabla\eta \cdot \mathbf{v})_0| + |(\boldsymbol{\eta} \cdot \mathbf{e}, \nabla\eta \cdot \mathbf{u})_0| + |(\boldsymbol{\eta} \cdot \mathbf{d}, \boldsymbol{\eta} \cdot \mathbf{e})_0| \\ &\leq |\boldsymbol{\eta} \cdot \mathbf{d}|_0|\nabla\eta \cdot \mathbf{v}|_0 + |\boldsymbol{\eta} \cdot \mathbf{e}|_0|\nabla\eta \cdot \mathbf{u}|_0 + |\boldsymbol{\eta} \cdot \mathbf{d}|_0|\boldsymbol{\eta} \cdot \mathbf{e}|_0. \end{aligned}$$

Now, by the inequality $(p+q)^2 \leq 2(p^2+q^2)$, it follows that

$$\begin{aligned} |\boldsymbol{\eta} \cdot \mathbf{d}|_0 &= |\eta d_1 + \eta d_2|_0 \\ &= \Big(\int_\Omega (\eta d_1 + \eta d_2)^2\, \mathrm{d}\mathbf{x} \Big)^{1/2} \\ &\leq \Big(2 \int_\Omega (\eta d_1)^2 + (\eta d_2)^2\, \mathrm{d}\mathbf{x} \Big)^{1/2} \\ &\leq \Big(2|\eta^2|_\infty \int_\Omega d_1^2 + d_2^2\, \mathrm{d}\mathbf{x} \Big)^{1/2} \\ &= (2|\eta^2|_\infty)^{1/2} |\mathbf{d}|_0. \end{aligned}$$

Similar reasoning for the other terms yields

$$\begin{aligned} |a_{\mathrm{M}}(\mathbf{u},\mathbf{v}) + a_{\mathrm{Q}}(\mathbf{u},\mathbf{v})| &\leq |\boldsymbol{\eta}\cdot\mathbf{d}|_0|\nabla\eta\cdot\mathbf{v}|_0 + |\boldsymbol{\eta}\cdot\mathbf{e}|_0|\nabla\eta\cdot\mathbf{u}|_0 + |\boldsymbol{\eta}\cdot\mathbf{d}|_0|\boldsymbol{\eta}\cdot\mathbf{e}|_0 \\ &\leq C_{\mathrm{ICE}}(|\mathbf{d}|_0|\mathbf{v}|_0 + |\mathbf{e}|_0|\mathbf{u}|_0 + |\mathbf{d}|_0|\mathbf{e}|_0), \end{aligned}$$

where $C_{\mathrm{ICE}} = C_{\mathrm{ICE}}(|\eta^2|_\infty, |\eta_x^2|_\infty, |\eta_y^2|_\infty)$. □

Lemma 5. *It holds that*

$$|\mathbf{d}|_0^2 \leq |\nabla u_1|_0^2 + |\nabla u_2|_0^2$$

and

$$|\mathbf{e}|_0^2 \leq |\nabla v_1|_0^2 + |\nabla v_2|_0^2$$

for all $\mathbf{u}, \mathbf{v} \in V$.

Proof. Let $\bar{\mathbf{d}} = [\partial u_1/\partial y, \partial u_2/\partial x]$. Then

$$\begin{aligned} |\mathbf{d}|_0^2 &= (\mathbf{d},\mathbf{d})_0 \\ &\leq (\mathbf{d},\mathbf{d})_0 + (\bar{\mathbf{d}},\bar{\mathbf{d}})_0 \\ &= (d_1,d_1)_0 + (d_2,d_2)_0 + (\bar{d_1},\bar{d_1})_0 + (\bar{d_2},\bar{d_2})_0 \\ &= (\nabla u_1, \nabla u_1)_0 + (\nabla u_2, \nabla u_2)_0 \\ &= |\nabla u_1|_0^2 + |\nabla u_2|_0^2 \end{aligned}$$

holds. Similar reasoning yields the second assertion. □

Lemma 6 (Linearform Continuity)**.** *The linearform* $f_{\mathrm{ICE}}(\mathbf{v})$ *is continuous.*

Proof. By definition of the two norms it is obvious that $|\mathbf{v}|_0 \leq |\mathbf{v}|_1$ for all $\mathbf{v} \in V$. Similarly, using Lemma 5, it is easy to see that $|\mathbf{e}|_0 \leq |\mathbf{v}|_1$ holds. Via the Cauchy-Schwarz inequality we get:

$$\begin{aligned} |f_{\mathrm{ICE}}(\mathbf{v})| &= |-2\int_\Omega \eta_t\nabla\eta\cdot\mathbf{v} + \eta\boldsymbol{\eta}_t\cdot\mathbf{e}\,\mathrm{d}\mathbf{x}| \\ &\leq 2|\int_\Omega \eta_t\nabla\eta\cdot\mathbf{v}\,\mathrm{d}\mathbf{x}| + |\int_\Omega \eta\boldsymbol{\eta}_t\cdot\mathbf{e}\,\mathrm{d}\mathbf{x}| \\ &\leq 2(|\eta_t\nabla\eta|_0|\mathbf{v}|_0 + |\eta\boldsymbol{\eta}_t|_0|\mathbf{e}|_0) \\ &\leq 2\cdot\max\{|\eta_t\nabla\eta|_0, |\eta\boldsymbol{\eta}_t|_0\}(|\mathbf{v}|_0 + |\mathbf{e}|_0) \\ &\leq 2\cdot\max\{|\eta_t\nabla\eta|_0, |\eta\boldsymbol{\eta}_t|_0\}|\mathbf{v}|_1 \end{aligned}$$

□

Lemma 7 (Bilinearform Continuity)**.** *The bilinearform $a_{\mathrm{ICE}}(\cdot,\cdot)$ is continuous.*

Proof. Following Schnörr (1991) we have

$$|a_{\mathrm{HS}}(\mathbf{u},\mathbf{v}) + a_{\mathrm{S}}(\mathbf{u},\mathbf{v})| \;\leq\; C_{\mathrm{HS}}(|\mathbf{u}|_0|\mathbf{v}|_0 + |\nabla u_1|_0|\nabla v_1|_0 + |\nabla u_2|_0|\nabla v_2|_0)$$

for some $C_{\mathrm{HS}} = C_{\mathrm{HS}}(\eta) > 0$. Lemma 4 then leads to

$$\begin{aligned}
|a_{\mathrm{ICE}}(\mathbf{u},\mathbf{v})| &\leq |a_{\mathrm{HS}}(\mathbf{u},\mathbf{v}) + a_{\mathrm{S}}(\mathbf{u},\mathbf{v})| + |a_{\mathrm{M}}(\mathbf{u},\mathbf{v}) + a_{\mathrm{Q}}(\mathbf{u},\mathbf{v})| \\
&\leq C_{\mathrm{HS}}(|\mathbf{u}|_0|\mathbf{v}|_0 + |\nabla u_1|_0|\nabla v_1|_0 + |\nabla u_2|_0|\nabla v_2|_0) \\
&\quad + C_{\mathrm{ICE}}(|\mathbf{d}|_0|\mathbf{v}|_0 + |\mathbf{e}|_0|\mathbf{u}|_0 + |\mathbf{d}|_0|\mathbf{e}|_0) \\
&= C(|\mathbf{u}|_0|\mathbf{v}|_0 + |\nabla u_1|_0|\nabla v_1|_0 + |\nabla u_2|_0|\nabla v_2|_0 \\
&\quad + |\mathbf{d}|_0|\mathbf{v}|_0 + |\mathbf{e}|_0|\mathbf{u}|_0 + |\mathbf{d}|_0|\mathbf{e}|_0)
\end{aligned}$$

and thus, by the Cauchy-Schwarz inequality,

$$\begin{aligned}
|a_{\mathrm{ICE}}(\mathbf{u},\mathbf{v})| \leq C \cdot (&|\mathbf{u}|_0^2 + |\nabla u_1|_0^2 + |\nabla u_2|_0^2 + |\mathbf{d}|_0^2 + |\mathbf{u}|_0^2 + |\mathbf{d}|_0^2)^{1/2} \\
&\cdot(|\mathbf{v}|_0^2 + |\nabla v_1|_0^2 + |\nabla v_2|_0^2 + |\mathbf{e}|_0^2 + |\mathbf{v}|_0^2 + |\mathbf{e}|_0^2)^{1/2}
\end{aligned}$$

for $C = \max\{C_{\mathrm{HS}}, C_{\mathrm{ICE}}\}$. Applying Lemma 5 then results in

$$\begin{aligned}
|a_{\mathrm{ICE}}(\mathbf{u},\mathbf{v})| &\leq C \cdot (2|\mathbf{u}|_0^2 + 3|\nabla u_1|_0^2 + 3|\nabla u_2|_0^2)^{1/2}(2|\mathbf{v}|_0^2 + 3|\nabla v_1|_0^2 + 3|\nabla v_2|_0^2)^{1/2} \\
&\leq \tilde{C}|\mathbf{u}|_1|\mathbf{v}|_1
\end{aligned}$$

with $\tilde{C} = 3C = 3^{1/2} \cdot 3^{1/2}C$ □

Lemma 8 (Convergence of Partial Derivative Norms)**.** *Consider a sequence $\mathbf{v}_n = (v_{1n}, v_{2n})$, $n \in \mathbb{N}$, in V such that $a_{\mathrm{ICE}}(\mathbf{v}_n,\mathbf{v}_n) \to 0$ as $n \to \infty$. Then the (squared) L^2 norms of the flow field partial derivatives converge to zero, i.e.:*

$$\lim_{n\to\infty} \int_\Omega \left(\frac{\partial v_{in}}{\partial z}\right)^2 \mathrm{d}\mathbf{x} = 0 \qquad \forall i \in \{1,2\}, z \in \{x,y\} \tag{4.36}$$

Proof. Recall that a_{ICE} is an integral over a sum of two squared and therefore nonnegative terms:

$$a_{\mathrm{ICE}}(\mathbf{v},\mathbf{v}) = \int_\Omega (\nabla\eta \cdot \mathbf{v} + \eta \cdot \mathbf{e})^2 \,\mathrm{d}\mathbf{x} + a_{\mathrm{S}}(\mathbf{v},\mathbf{v}).$$

Hence, if $a_{\mathrm{ICE}}(\mathbf{v}_n,\mathbf{v}_n) \to 0$ it follows that $a_{\mathrm{S}}(\mathbf{v}_n,\mathbf{v}_n) \to 0$. Furthermore, since

$$a_{\mathrm{S}}(\mathbf{v},\mathbf{v}) = \int_\Omega \alpha^2(v_{1x}^2 + v_{1y}^2 + v_{2x}^2 + v_{2y}^2) \,\mathrm{d}\mathbf{x},$$

the same argument allows to conclude that Equation (4.36) holds. □

Lemma 9 (Boundedness of Flow Field Component Norms)**.** *Consider a sequence* $\mathbf{v}_n = (v_{1n}, v_{2n})$*,* $n \in \mathbb{N}$*, in* V *with* $|\mathbf{v}_n|_V = 1$*. Then the (squared)* L^2 *norms of the flow field components are bounded, i.e.:*

$$0 \leq \int_\Omega v_{in}^2 \, \mathrm{d}\mathbf{x} \leq 1 \qquad \forall i \in \{1, 2\}.$$

Proof. The first inequality follows directly from the nonnegativity of the integrand. Now recall that

$$|\mathbf{v}_n|_V = 1 \quad = \quad \int_\Omega v_{1n}^2 + v_{2n}^2 \, \mathrm{d}\mathbf{x} + \int_\Omega |\nabla v_{10n}|^2 \, \mathrm{d}\mathbf{x} + \int_\Omega |\nabla v_{20n}|^2 \, \mathrm{d}\mathbf{x}$$

Hence, since all summands are equal or greater than zero, it follows that

$$\int_\Omega v_{1n}^2 \, \mathrm{d}\mathbf{x} = 1 - (\int_\Omega v_{2n}^2 \, \mathrm{d}\mathbf{x} + \int_\Omega |\nabla v_{10n}|^2 \, \mathrm{d}\mathbf{x} + \int_\Omega |\nabla v_{20n}|^2 \, \mathrm{d}\mathbf{x}) \leq 1,$$

as well as

$$\int_\Omega v_{2n}^2 \, \mathrm{d}\mathbf{x} = 1 - (\int_\Omega v_{1n}^2 \, \mathrm{d}\mathbf{x} + \int_\Omega |\nabla v_{10n}|^2 \, \mathrm{d}\mathbf{x} + \int_\Omega |\nabla v_{20n}|^2 \, \mathrm{d}\mathbf{x}) \leq 1.$$

□

Lemma 10. *Let* r_i *and* s_i *with* $i \in \mathcal{I}_C$ *two finite sets of* $|\cdot|_\infty$*-bounded functions on the image domain* Ω *and consider two sequences of corresponding sets of sequences of functions* p_{in} *and* q_{in} *with* $n \in \mathbb{N}$*. If for some constants* C_1 *and* C_2

$$C_1 \leq \int_\Omega p_{in}^2 \, \mathrm{d}\mathbf{x} \leq C_2 \quad \forall n$$

for fixed $i \in \mathcal{I}_C$ *and* $\lim_{n\to\infty} \int_\Omega q_n^2 \, \mathrm{d}\mathbf{x} = 0$ *then it holds that*

$$\lim_{n\to\infty} \Big| \int_\Omega \sum_i (r_i p_{in})(s_i q_{in}) \, \mathrm{d}\mathbf{x} \Big| = 0$$

Proof. We apply the triangle inequality for sums and integrals as well as the Cauchy-Schwarz inequality for integrals as follows:

$$\begin{aligned}
\Big| \int_\Omega \sum_i (r_i p_{in})(s_i q_{in}) \, \mathrm{d}\mathbf{x} \Big| &\leq \sum_i \int_\Omega \Big| (r_i p_{in})(s_i q_{in}) \Big| \, \mathrm{d}\mathbf{x} \\
&\leq \sum_i \sqrt{\Big(\int_\Omega r_i^2 p_{in}^2 \, \mathrm{d}\mathbf{x} \Big) \Big(\int_\Omega s_i^2 q_{in}^2 \, \mathrm{d}\mathbf{x} \Big)} \\
&\leq \sum_i \sqrt{|r_i^2|_\infty |s_i^2|_\infty \Big(\underbrace{\int_\Omega p_{in}^2 \, \mathrm{d}\mathbf{x}}_{t_1} \Big) \Big(\underbrace{\int_\Omega q_{in}^2 \, \mathrm{d}\mathbf{x}}_{t_2} \Big)}.
\end{aligned}$$

Now, since for every $i \in \mathcal{I}_C$ the term t_1 is bounded and t_2 converges to zero the assertion follows. □

Lemma 11 (V-ellipticity)**.** *The bilinearform a_{ICE} is V-elliptic.*

Proof. As shown by Schnörr (1991), $a_{HS} + a_S$ is V-elliptic under the assumptions of Theorem 8. Now assume that a_{ICE} is not V-elliptic. We will show that this contradicts the V-ellipticity of $a_{HS} + a_S$.

If a_{ICE} is not V-elliptic then there exists a sequence $\mathbf{v}_n$ in V with $|\mathbf{v}_n|_1 = 1$ for all n and $a_{ICE}(\mathbf{v}_n, \mathbf{v}_n) \to 0$ for $n \to \infty$. Recall that

$$a_{\mathrm{ICE}}(\mathbf{v}, \mathbf{v}) = a_{\mathrm{HS}}(\mathbf{v}, \mathbf{v}) + a_{\mathrm{M}}(\mathbf{v}, \mathbf{v}) + a_{\mathrm{Q}}(\mathbf{v}, \mathbf{v}) + a_{\mathrm{S}}(\mathbf{v}, \mathbf{v})$$

and

$$\begin{aligned} a_{\mathrm{M}}(\mathbf{v}, \mathbf{v}) &= 2 \int_\Omega \eta \Big[\eta_x v_{1,x} v_1 + \eta_y v_{1x} v_2 + \eta_x v_{2,y} v_1 + \eta_y v_2 v_{2y} \Big] \, \mathrm{d}\mathbf{x}, \\ a_{\mathrm{Q}}(\mathbf{v}, \mathbf{v}) &= \int_\Omega \eta^2 \Big[v_{1x}^2 + 2 v_{1x} v_{2y} + v_{2y}^2 \Big] \, \mathrm{d}\mathbf{x}. \end{aligned}$$

Consider the structure of the summands, say t_i, of $a_{\mathrm{M}}(\mathbf{v}_n, \mathbf{v}_n)$ and $a_{\mathrm{Q}}(\mathbf{v}_n, \mathbf{v}_n)$ for a given n:

$$t_i = c_i \cdot r_i p_{in} \cdot s_i q_{in}, \quad i \in \mathcal{I}_C,$$

where $c_i \in \{1, 2\}$, $r_i, s_i \in \{\eta, \eta_x, \eta_y\}$ and the flow field components and their partial derivatives are represented by p_{in} and q_{in}. We now yield to apply Lemma 10 to these terms and thus to derive convergence properties for $a_{\mathrm{M}}(\mathbf{v}_n, \mathbf{v}_n)$ and $a_{\mathrm{Q}}(\mathbf{v}_n, \mathbf{v}_n)$.

Firstly, we note that, as a prerequisite, the coefficient functions r_i and s_i are bounded with respect to $|\cdot|_\infty$, i.e.

$$r_i, s_i \in L^\infty. \tag{4.37}$$

Second, each of the terms t_i involves at least one occurrence of a partial derivative of a flow field component, say q_{in}. Now, due to the assumption that $a_{ICE}(\mathbf{v}_n, \mathbf{v}_n) \to 0$ for $n \to \infty$, Lemma 8 is applicable. Hence,

$$\int_\Omega q_{in}^2 \mathrm{d}\mathbf{x} \to 0. \tag{4.38}$$

The last factor, say p_{in}, is either a partial derivative or a flow field component. In case of a partial derivative, the same argument as for the q_{in} applies and p_{in} is L^2-bounded due to the convergence with respect to the L^2 norm. If p_{in} is a flow fields component, Lemma 9 is applicable. Hence,

$$C_{i1} \leq p_{in} \leq C_{i2}, \tag{4.39}$$

for some $C_{ij} > 0 \in \mathbb{R}$. Equation (4.37) to (4.39) now allow the application of Lemma 10 to the the sequences $a_{\mathrm{M}}(\mathbf{v}_n, \mathbf{v}_n)$ and $a_{\mathrm{Q}}(\mathbf{v}_n, \mathbf{v}_n)$. Hence,

$$\lim_{n \to \infty} a_{\mathrm{M}}(\mathbf{v}_n, \mathbf{v}_n) = 0$$

and

$$\lim_{n\to\infty} a_{\mathrm{Q}}(\mathbf{v}_n, \mathbf{v}_n) = 0.$$

Finally, recall the definition of a_{HS} and a_{S} and observe that

$$\begin{aligned} |a_{\mathrm{HS}} + a_{\mathrm{S}}| &= |a_{\mathrm{HS}} + a_{\mathrm{S}} - a_{\mathrm{ICE}} + a_{\mathrm{ICE}}| \\ &\leq |a_{\mathrm{M}}| + |a_{\mathrm{Q}}| + |a_{\mathrm{ICE}}|, \end{aligned}$$

where we omitted the arguments $(\mathbf{v}_n, \mathbf{v}_n)$. Hence, one has that $a_{\mathrm{HS}}(v_n, v_n) \to 0$ for $n \to 0$. This leads to a contradiction to the fact that $a_{\mathrm{HS}} + a_{\mathrm{S}}$ is V-elliptic in view of the assumption $|\mathbf{v}|_1 = 1$. □

Proof of Theorem 8. The previous lemmata lead to the following assertions:

- The linearform f_{ICE} is continuous (Lemma 6)
- The bilinearform a_{ICE} is continuous (Lemma 7)
- The bilinearform a_{ICE} is V-elliptic (Lemma 11).

Hence, the Lemma of Lax-Milgram (Lemma 3 in Section 4.1.2) is applicable to the rewritten functional

$$\mathcal{J}_{\mathrm{ICE}}(\mathbf{v}) = a_{\mathrm{ICE}}(\mathbf{v}, \mathbf{v}) + f_{\mathrm{ICE}}(\mathbf{v}, \mathbf{v}) + \mathrm{const.}$$

and a unique minimizer of $\mathcal{J}_{\mathrm{ICE}}(\mathbf{v})$ with a well-posed discretization exists. □

4.5.3 The Euler-Lagrange Equations of Compressible Flow

We will now derive the Euler-Lagrange equations (see Section 4.1.2, Equation (4.8)) of the compressible flow functional given in Equation (4.34) with respect to the flow variables. Let, as before,

$$\mathcal{J}_{\mathrm{ICE}}(\mathbf{w}) = \int_\Omega (\mathbf{w} \cdot \nabla\eta + \eta_t + \eta \operatorname{div}(\mathbf{w}))^2 + \alpha^2(|\nabla w_1|^2 + |\nabla w_2|^2)\, \mathrm{d}\mathbf{x}.$$

We compute

$$\begin{aligned} \mathcal{J}_{\mathrm{ICE}}(\mathbf{u} + \tau\mathbf{v}) = &\int_\Omega ((\mathbf{u} + \tau\mathbf{v}) \cdot \nabla\eta + \eta_t + \eta \operatorname{div}(\mathbf{u} + \tau\mathbf{v}))^2\, \mathrm{d}\mathbf{x} \\ &+ \int_\Omega \alpha^2(|\nabla(u_1 + \tau v_1)|^2 + |\nabla(u_2 + \tau v_2)|^2)\, \mathrm{d}\mathbf{x}, \end{aligned}$$

hence

$$\begin{aligned}\frac{d}{d\tau}\mathcal{J}_{\mathrm{ICE}}(\mathbf{u}+\tau\mathbf{v}) &= \frac{d}{d\tau}\int_\Omega ((\mathbf{u}+\tau\mathbf{v})\cdot\nabla\eta+\eta_t+\eta\,\mathrm{div}(\mathbf{u}+\tau\mathbf{v}))^2\,\mathrm{d}\mathbf{x}\\ &\quad+\frac{d}{d\tau}\int_\Omega \alpha^2(|\nabla(u_1+\tau v_1)|^2+|\nabla(u_2+\tau v_2)|^2)\,\mathrm{d}\mathbf{x}\\ &= 2\int_\Omega ((\mathbf{u}+\tau\mathbf{v})\cdot\nabla\eta+\eta_t+\eta\,\mathrm{div}(\mathbf{u}+\tau\mathbf{v}))(\mathbf{v}\cdot\nabla\eta+\eta\,\mathrm{div}(\mathbf{v}))\,\mathrm{d}\mathbf{x}\\ &\quad+2\int_\Omega \alpha^2(\nabla u_1\cdot\nabla v_1+\tau|\nabla v_1|^2+\nabla u_2\cdot\nabla v_2+\tau|\nabla v_2|^2)\,\mathrm{d}\mathbf{x}.\end{aligned}$$

The first variation is given by:

$$\begin{aligned}\frac{1}{2}\frac{d}{d\tau}\mathcal{J}_{\mathrm{ICE}}(\mathbf{u}+\tau\mathbf{v})|_{\tau=0} &= \int_\Omega[(\mathbf{u}\cdot\nabla\eta+\eta_t+\eta\,\mathrm{div}(\mathbf{u}))\nabla\eta]\cdot\mathbf{v}\,\mathrm{d}\mathbf{x}\\ &\quad+\int_\Omega[(\mathbf{u}\cdot\nabla\eta+\eta_t+\eta\,\mathrm{div}(\mathbf{u}))\eta]\,\mathrm{div}(\mathbf{v})\,\mathrm{d}\mathbf{x}\\ &\quad+\int_\Omega\alpha^2(\nabla u_1\cdot\nabla v_1+\nabla u_2\cdot\nabla v_2)\,\mathrm{d}\mathbf{x}\end{aligned}$$

Applying Green's formula (n is the outer normal vector at the boundary $\partial\Omega$),

$$\int_\Omega\nabla u_1\cdot\nabla v_1=-\int_\Omega v_1\Delta u_1\,\mathrm{d}\mathbf{x}+\int_{\partial\Omega}\frac{\partial u_1}{\partial n}v_1\,\mathrm{d}\mathbf{x}$$

and

$$\int_\Omega\nabla u_2\cdot\nabla v_2=-\int_\Omega v_2\Delta u_2\,\mathrm{d}\mathbf{x}+\int_{\partial\Omega}\frac{\partial u_2}{\partial n}v_2\,\mathrm{d}\mathbf{x},$$

we obtain

$$\begin{aligned}\frac{1}{2}\frac{d}{d\tau}\mathcal{J}_{\mathrm{ICE}}(\mathbf{u}+\tau\mathbf{v})|_{\tau=0} &= \int_\Omega[(\mathbf{u}\cdot\nabla\eta+\eta_t+\eta\,\mathrm{div}(\mathbf{u}))\nabla\eta]\cdot\mathbf{v}\,\mathrm{d}\mathbf{x}\\ &\quad+\int_\Omega[(\mathbf{u}\cdot\nabla\eta+\eta_t+\eta\,\mathrm{div}(\mathbf{u}))\eta]\,\mathrm{div}(\mathbf{v})\,\mathrm{d}\mathbf{x}\\ &\quad-\int_\Omega\alpha^2(v_1\Delta u_1+v_2\Delta u_2)\,\mathrm{d}\mathbf{x}\\ &\quad+\int_{\partial\Omega}\frac{\partial u_1}{\partial n}v_1+\frac{\partial u_2}{\partial n}v_2\,\mathrm{d}\mathbf{x}.\end{aligned}$$

Partial integration of the second term in this equation yields

$$\begin{aligned}\frac{1}{2}\frac{d}{d\tau}\mathcal{J}_{\mathrm{ICE}}(\mathbf{u}+\tau\mathbf{v})|_{\tau=0} &= \int_\Omega [(\mathbf{u}\cdot\nabla\eta+\eta_t+\eta\,\mathrm{div}(\mathbf{u}))\nabla\eta]\cdot\mathbf{v}\;\mathrm{d}\mathbf{x}\\ &\quad -\int_\Omega \mathbf{v}\cdot\nabla[(\mathbf{u}\cdot\nabla\eta+\eta_t+\eta\,\mathrm{div}(\mathbf{u}))\eta]\;\mathrm{d}\mathbf{x}\\ &\quad +\int_{\partial\Omega}[(\mathbf{u}\cdot\nabla\eta+\eta_t+\eta\,\mathrm{div}(\mathbf{u}))\eta]\mathbf{v}\cdot n\;\mathrm{d}\mathbf{S}\\ &\quad -\int_\Omega \alpha^2(v_1\Delta u_1+v_2\Delta u_2)\;\mathrm{d}\mathbf{x}\\ &\quad +\int_{\partial\Omega}\frac{\partial u_1}{\partial n}v_1+\frac{\partial u_2}{\partial n}v_2\;\mathrm{d}\mathbf{x}.\end{aligned}$$

Here, the result of applying the product rule to the second term cancels out the first term. Hence

$$\begin{aligned}\frac{1}{2}\frac{d}{d\tau}\mathcal{J}_{\mathrm{ICE}}(\mathbf{u}+\tau\mathbf{v})|_{\tau=0} &= -\int_\Omega \mathbf{v}\cdot\Big[\eta\nabla[(\mathbf{u}\cdot\nabla\eta+\eta_t+\eta\,\mathrm{div}(\mathbf{u}))]\Big]\;\mathrm{d}\mathbf{x}\\ &\quad +\int_{\partial\Omega}[(\mathbf{u}\cdot\nabla\eta+\eta_t+\eta\,\mathrm{div}(\mathbf{u}))\eta]\mathbf{v}\cdot n\;\mathrm{d}\mathbf{S}\\ &\quad -\int_\Omega \alpha^2(v_1\Delta u_1+v_2\Delta u_2)\;\mathrm{d}\mathbf{x}\\ &\quad +\int_{\partial\Omega}\frac{\partial u_1}{\partial n}v_1+\frac{\partial u_2}{\partial n}v_2\;\mathrm{d}\mathbf{x}.\end{aligned}$$

If $\mathbf{u}$ is a minimizer of this functional, this expression has to vanish for all functions $\mathbf{v}\in V$. Thus, we obtain

$$\begin{aligned}0 &= \eta\frac{\partial}{\partial x}(\mathbf{u}\nabla\eta+\eta_t+\eta\,\mathrm{div}(\mathbf{u}))+\alpha^2\Delta u_1,\\ 0 &= \eta\frac{\partial}{\partial y}(\mathbf{u}\nabla\eta+\eta_t+\eta\,\mathrm{div}(\mathbf{u}))+\alpha^2\Delta u_2,\\ 0 &= \frac{\partial u_1}{\partial n}+n_1\eta[\mathbf{u}\nabla\eta+\eta_t+\eta\,\mathrm{div}(\mathbf{u})]\text{ on }\partial\Omega,\\ 0 &= \frac{\partial u_2}{\partial n}+n_2\eta[\mathbf{u}\nabla\eta+\eta_t+\eta\,\mathrm{div}(\mathbf{u})]\text{ on }\partial\Omega.\end{aligned}$$

Hence, under the assumption of vanishing material density on the boundary, one has the Euler-Lagrange equations

$$\begin{aligned}0 &= \eta\frac{\partial}{\partial x}(\mathbf{u}\nabla\eta+\eta_t+\eta\,\mathrm{div}(\mathbf{u}))+\alpha^2\Delta u_1,\quad 0=\frac{\partial u_1}{\partial n}\text{ on }\partial\Omega\\ 0 &= \eta\frac{\partial}{\partial y}(\mathbf{u}\nabla\eta+\eta_t+\eta\,\mathrm{div}(\mathbf{u}))+\alpha^2\Delta u_2,\quad 0=\frac{\partial u_2}{\partial n}\text{ on }\partial\Omega.\end{aligned}$$

4.5.4 The Euler-Lagrange Equations of Material Densities

In the following we will derive the Euler-Lagrange equation for both, the the Horn & Schunck flow model (see Section 4.1.3) and the compressible flow model ((see Section 4.2.1) with respect to the material density. Furthermore, we will point out connections to the class of SPDEs elaborated on in Section 4.1.4.

Incompressible Flow

The smoothness term of the flow field is ignored here as it does not depend on the material density and thus does not enter the Euler-Lagrange equations. Hence, we are left with the Horn & Schunck data term

$$L = L_{\mathrm{HS}}(\eta) = (\nabla\eta \cdot \mathbf{w} + \eta_t)^2.$$

From this it follows that

$$\begin{aligned} L_\eta &= 0 \\ \tfrac{1}{2}L_{\eta_x} &= v_1(\nabla\eta \cdot \mathbf{w} + \eta_t) = v_1(\mathbf{w}^T\nabla\eta) + v_1\eta_t, \\ \tfrac{1}{2}L_{\eta_y} &= v_2(\nabla\eta \cdot \mathbf{w} + \eta_t) = v_2(\mathbf{w}^T\nabla\eta) + v_2\eta_t, \\ \tfrac{1}{2}L_{\eta_t} &= 1(\nabla\eta \cdot \mathbf{w} + \eta_t) = 1(\mathbf{w}^T\nabla\eta) + 1\eta_t. \end{aligned}$$

The Euler-Lagrange equation is thus

$$0 = (\nabla\eta \cdot \mathbf{w} + \eta_t)(\nabla \cdot \mathbf{w}) + \mathbf{w}^T(\nabla\eta_t + H_\eta\mathbf{w} + J_{\mathbf{w}}^T\nabla\eta). \tag{4.40}$$

where H_η is the Hessian of η and $J_{\mathbf{w}}$ is the Jacobi matrix of $\mathbf{w}$. Alternatively, one can write the vector $[L_{\eta_x}, L_{\eta_y}, L_{\eta_t}]^T$ using $\tilde{\mathbf{w}} = [v_1, v_2, 1]^T$ as

$$\tfrac{1}{2}[L_{\eta_x}, L_{\eta_y}, L_{\eta_t}]^T = \tilde{\mathbf{w}}\tilde{\mathbf{w}}^T\tilde{\nabla}\eta,$$

where $\tilde{\nabla} = [\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial t}]$ denotes the gradient including the partial derivative along time. The Laplacian of this expression, i.e. the right hand side of the Euler-Lagrange equation then simplifies and results in

$$0 = \nabla \cdot [L_{\eta_x}, L_{\eta_y}, L_{\eta_t}]^T = \tilde{\nabla} \cdot (\tilde{\mathbf{w}}\tilde{\mathbf{w}}^T\tilde{\nabla}\eta) = (\tilde{\nabla}^T\tilde{\mathbf{w}}\tilde{\mathbf{w}}^T\tilde{\nabla})\eta. \tag{4.41}$$

Now consider the common situation of applying the Horn and Schunck algorithm to a pair of consecutive images η and $\bar{\eta}$ where the partial differential w.r.t. time is estimated by the forward difference $\bar{\eta} - \eta$. Equation (4.41) then reads as

$$0 = \nabla^T(\mathbf{w}(\eta - \bar{\eta})) + (\nabla^T\mathbf{w}\mathbf{w}^T\nabla)\eta,$$

which holds if and only if

$$\nabla(\mathbf{w}\bar{\eta}) \quad = \quad (\nabla^T\mathbf{w} + \nabla^T\mathbf{w}\mathbf{w}^T\nabla)\eta. \tag{4.42}$$

This equation, in particular the homogeneous part, seems remarkably similar to the SPDE described by Equation (4.27) in Section 4.1.4. However, an in-depth analysis of this connection has to be left to future research. Lastly, we would like to note that when assuming incompressibility (and thus zero divergence) as well as temporal constancy of the flow field (in between the two image frames), Equation (4.42) reduces to

$$\mathbf{w}^T\nabla\bar{\eta} \quad = \quad \mathbf{w}^T\nabla\eta + (\nabla^T\mathbf{w}\mathbf{w}^T\nabla)\eta. \tag{4.43}$$

Compressible Flow

We will now turn to the case of compressible flow described in Section 4.4.1. The respective data term

$$\begin{aligned} L = L_{\mathrm{ICE}}(\eta) &= (\tilde{\mathbf{w}}^T\nabla\eta + \eta\nabla^T\tilde{\mathbf{w}})^2 \\ &= (\tilde{\mathbf{w}}^T\nabla\eta)^2 + 2(\tilde{\mathbf{w}}^T\nabla\eta)(\eta\,\mathrm{div}(\tilde{\mathbf{w}})) + \eta^2\,\mathrm{div}(\tilde{\mathbf{w}})^2 \end{aligned}$$

results in the partial derivatives

$$\begin{aligned} \tfrac{1}{2}L_\eta &= (\tilde{\mathbf{w}}^T\nabla\eta)\,\mathrm{div}(\tilde{\mathbf{w}}) + \eta\,\mathrm{div}(\tilde{\mathbf{w}})^2 = \mathrm{div}(\tilde{\mathbf{w}})(\mathrm{div}(\eta\tilde{\mathbf{w}})), \\ \tfrac{1}{2}L_{\eta_x} &= v_1(\tilde{\mathbf{w}}^T\nabla\eta) + v_1(\eta\,\mathrm{div}(\tilde{\mathbf{w}})), \\ \tfrac{1}{2}L_{\eta_y} &= v_2(\tilde{\mathbf{w}}^T\nabla\eta) + v_2(\eta\,\mathrm{div}(\tilde{\mathbf{w}})), \\ \tfrac{1}{2}L_{\eta_t} &= 1(\tilde{\mathbf{w}}^T\nabla\eta) + 1(\eta\,\mathrm{div}(\tilde{\mathbf{w}})), \end{aligned}$$

and thus

$$\tfrac{1}{2}[L_{\eta_x}, L_{\eta_y}, L_{\eta_t}] \quad = \quad \tilde{\mathbf{w}}[(\tilde{\mathbf{w}}^T\nabla\eta) + (\eta\,\mathrm{div}(\tilde{\mathbf{w}}))] = \tilde{\mathbf{w}}(\mathrm{div}(\eta\tilde{\mathbf{w}})).$$

This leads to

$$\begin{aligned} 0 &= \mathrm{div}(\tilde{\mathbf{w}})(\mathrm{div}(\eta\tilde{\mathbf{w}})) - \nabla^T(\tilde{\mathbf{w}}(\mathrm{div}(\eta\tilde{\mathbf{w}}))) \\ &= \mathrm{div}(\tilde{\mathbf{w}})\tilde{q} - \nabla^T(\tilde{\mathbf{w}}\tilde{q}) \\ &= \mathrm{div}(\tilde{\mathbf{w}})\tilde{q} - [\mathrm{div}(\tilde{\mathbf{w}}) + \tilde{\mathbf{w}}^T\nabla\tilde{q}] \\ &= -\tilde{\mathbf{w}}^T\nabla\tilde{q} \\ &= -\tilde{\mathbf{w}}^T\nabla(\nabla^T(\tilde{\mathbf{w}}\eta)) \\ &= -\tilde{\mathbf{w}}^T\nabla\nabla^T\tilde{\mathbf{w}}\eta \end{aligned}$$

where we introduced the abbreviation $\tilde{q} = \mathrm{div}(\tilde{\mathbf{w}}\eta)$. This expression can easily be transformed to

$$0 \quad = \quad (\nabla^T \tilde{\mathbf{w}} \tilde{\mathbf{w}}^T \nabla)\eta + \eta(\tilde{\mathbf{w}}^T \nabla \nabla^T \tilde{\mathbf{w}}) \tag{4.44}$$

Note due to the notation we chose differentials act on every multiplier to their right. The second term does therefore not contain a differential expression in g.

If we now focus on to consecutive image frames, we obtain

$$\nabla(\mathbf{w}\bar{\eta}) \quad = \quad (\nabla^T \mathbf{w}\mathbf{w}^T \nabla + \nabla^T \mathbf{w})\eta + \eta(\mathbf{w}^T \nabla \nabla^T \mathbf{w}).$$

By using the notation $z = -\mathbf{w}^T \nabla \nabla^T \mathbf{w}$ this can be written as

$$z^{-1}\nabla(\mathbf{w}\bar{\eta}) \quad = \quad (1 - z^{-1}\nabla^T \mathbf{w} - z^{-1}\nabla^T \mathbf{w}\mathbf{w}^T \nabla)\eta. \tag{4.45}$$

Comparing this result to the SPDE in Equation (4.27) it is clear that further investigation regarding this topic might be fruitful. In particular, establishing a clear connection between the deformation determinant $\det(F)$ in Equation (4.27) and z^{-1} appears to be an important first step.

5 Conclusion

The thesis at hand presents statistical methodology and theoretical work related to the remote assessment of atmospheric aerosol occurrence and dynamics. Naturally, this subject lives at the intersection of physical models, statistical analysis and mathematical image processing. We develop a Bayesian framework that consistently reflects these perspectives and leverages the cutting-edge INLA inference technique to detect dust storms and consecutively capture the respective atmospheric transport process. In doing so, the latent processes we investigate are inherently endowed with uncertainty quantification in terms of the respective marginal posterior distributions. Moreover, hyper parameters are integrated over and by evading computationally expensive MCMC procedures we develop algorithms that lend themselves towards real time analysis of atmospheric processes.

The components of both, the detection and the motion estimation models, are structured as Bayesian hierarchical models with generalized linear models as top stages and Gaussian Markov random fields representing the latent variables. We elucidate the principles of these graphical models as well as the INLA inference approach in Chapter 2. Their intrinsic modularity, favorable scaling properties in prospect of big data and recently increasing scientific attention suggest the potential of adapting our methods to a wide range of applications.

At the very heart the presented framework evolves around the continuity equation for compressible fluid flow. This equation physically ties together the dust density and its flow field. In Chapter 3 a data driven approximation of this density is developed via a latent predictor steering a binomial likelihood that expresses the dichotomy of dusty versus pristine spatio-temporal locations. The resulting method of latent signal mapping successfully suppresses measurement artifacts not related to dust activity and, at the same time, reliably detects dust throughout the forenoon by adapting to radiative surface properties of the Earth. However, considerable work remains. With respect to the physical interpretation of the dust predictor multiple extensions of our methodology are conceivable. Aerosol optical depth measurements could enter the likelihood of the LSM method as a replacement or as an addition to binomial labels over data labeled by experts. For example, as a direct exponential function of the latent field (when omitting the sign), a Laplacian likelihood could be employed to relate the transmissivity of SEVIRI measurements to the optical depth. This way, additional assumptions on the grain size distribution of the dust allow to estimate the dust density. Providing the latent part of LSM with scientific attention is promising

as well. At the current stage, time as a covariate is not part of the model even though the framework's flexibility clearly allows otherwise. Separating the effect of time on the SFI appearance from radiative surface properties and other constituents (e.g. satellite viewing angle) might leverage the method to robust dust detection even at night times. As a mode of validation and further gain of physical interpretability this separation also suggests a comparison to or an integration of results from atmospheric simulation studies like that of Brindley et al. (2012). Lastly, our current approach does not impose any spatial or temporal smoothness assumptions with respect to the dust predictor. A straight forward and frequently applied way to do so is to couple the local predictor with its spatial or temporal grid neighbors, e.g. via a CAR model as an intermediate layer between likelihood and latent functions over the covariates. In the long run, however, directly integrating an estimate of the flow as consistency criterion appears to be more promising.

Chapter 4 is concerned with the second kind of unknowns of the continuity equation, the estimation of a flow field given an approximation to the material density. By employing a latent Gaussian Markov random field for the transition between the continuous formulation and a discretized representation an important link between variational methods and classical literature on generalized linear models is established. Traditionally, inferring the Horn and Schunck optical flow was subject to solving a variational formulation of the problem via the corresponding Euler-Lagrange equations. The equivalence to a formulation as a GMRF and the extension to compressible fluid dynamics provides considerable applied advances that can now be pursued. This relates to projecting the dust storm into the future, as well as "rewinding" the storm to pinpoint its source. The advantage of our statistical approach is that it inherently enables the uncertainty of such assessments to be expressed. This, in turn, will allow us to issue probabilistic forecasts and leverage the recent work in forecasting methodology (Gneiting and Raftery, 2005; Schefzik et al., 2013). Such probabilistic forecasts would be of considerable interest to the Earth observation community and could also be fed into larger models of global transport phenomena.

The modularity of our framework and the flexibility of the INLA method suggest the investigation of several extensions to the flow estimation technique. Due to the demand for source estimators adding a respective variable to the continuity equation appears to be useful. Unless specified otherwise this variable would implicitly capture the vertical components of the three-dimensional counterpart of the equation and could hence capture the injection of soil from the ground into the atmosphere. Yet, this approach would require further characterization of the injection process as the system still remains under-determined and prevents a separation of the respective random effects. With respect to current developments in the image processing community our probabilistic inference paradigm is of two-fold interest. Firstly, likelihood functions that are more robust against outliers than the Gaussian distribution are a common mode to tackle the problem of signal intrinsic noise. Respective changes to the INLA algorithm are readily implemented and would provide a powerful tool for in-depth

stochasticity analyses. Secondly, motion estimation for imagery of rigid bodies is a vividly discussed subject and requires to allow for discontinuities in the flow field. In this context (see, e.g. Roth and Black (2005)), heavy-tailed distributions like the Laplacian have been shown to be more suitable to model flow derivatives than Gaussians. At the same time, the Laplace distribution can be interpreted as a normal variance mixture model or a scale mixture of Gaussians (Fahrmeir and Tutz, 2001). As shown by Eltoft et al. (2006), this can be generalized to multivariate Laplacians being generated by a multivariate Gaussian integrated over a univariate exponential scale prior. Related Bayesian approaches have recently been proposed by Bolin (2013) and Bolin et al. (2014) in context of different environmental and medical applications. Both, robust likelihood functions and discontinuity preserving latent fields are promising methodological candidates to alleviate the underestimation of flow velocity we observed in our forecast experiments. In particular the latter would reduce the effect of flow field smoothing at the often very prominent front of dust plumes.

An investigation of the variational perspective on the continuity equation for compressible flow that concludes Chapter 4 leads to further insight with respect to the current nature of the model as well as future research. We extend results of Schnörr (1991) for incompressible flow by providing a proof for existence and uniqueness of (continuous) compressible flow under mild restrictions to an underlying Sobolev space. As a corollary, the emerging discrete quadratic minimization problem is well posed and its GMRF counterpart is shown to have a symmetric and positive definite precision matrix. Lastly, we provide first steps towards an inversion of the inference process, i.e. estimating the development of a dust density from a given flow field. The Euler-Lagrange equations arising from this minimization problem suggest to be solved for readily although respective analytical guarantees remain to be investigated. It is, however, appealing that work like that of Lindgren et al. (2011) as well as Simpson et al. (2012) reveals links between Gaussian fields and GMRFs via SPDEs. In particular, further in-depth mathematical analysis of the connection between these SPDEs and the aforementioned Euler-Lagrange equations would provide an extensive probabilistic basis for the inference inversion. In doing so, a major cornerstone would be laid with respect to a long term perspective of space-born dust assessment since the SPDE approach can be seen as a natural extension of Matérn fields to smooth manifolds like the sphere. Having both conditional models formulated properly, i.e. the dust density given its flow and vice versa, alternating estimation schemes like expectation maximization can be applied. Such an estimation of the joint distribution and respective convergence analysis would then profit from assertions about the solution characteristics of the conditional distributions.

Bibliography

S. Aberg, F. Lindgren, A. Malmberg, J. Holst, and U. Holst. An image warping approach to spatio-temporal modelling. *Environmetrics*, 16(8):833–848, December 2005. ISSN 1180-4009. doi: 10.1002/env.741.

S. Alonso-Pérez, E. Cuevas, X. Querol, J. C. Guerra, and C. Prez. African dust source regions for observed dust outbreaks over the Subtropical Eastern North Atlantic region, above 25°N. *Journal of Arid Environments*, 78(0):100–109, March 2012. ISSN 0140-1963. doi: 10.1016/j.jaridenv.2011.11.013.

Ian Ashpole and Richard Washington. An automated dust detection using SEVIRI: A multiyear climatology of summertime dustiness in the Central and Western Sahara. *Journal of Geophysical Research: Atmospheres*, 117(D8):D08202, April 2012. ISSN 2156-2202. doi: 10.1029/2011JD016845.

Fabian E. Bachl and Christoph S. Garbe. Classifying and tracking dust plumes from passive remote sensing. In L. Ouwehand, editor, *Proceedings of the ESA, SOLAS & EGU Joint Conference 'Earth Observation for Ocean-Atmosphere Interaction Science'*, volume 703 of *ESA Special Publication*, pages S1–3, Frascati, Italy, March 2012. European Space Agency, European Space Agency Communications.

Fabian E. Bachl, P. Fieguth, and Christoph S. Garbe. A bayesian approach to spaceborn hyperspectral optical flow estimation on dust aerosols. In *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*, pages 256–259. IEEE, July 2012. ISBN 978-1-4673-1160-1. doi: 10.1109/IGARSS.2012.6351589.

Fabian E. Bachl, P. Fieguth, and Christoph S. Garbe. Bayesian inference on integrated continuity fluid flows and their application to dust aerosols. In *Geoscience and Remote Sensing Symposium (IGARSS), 2013 IEEE International*, pages 2246–2249. IEEE, July 2013a. ISBN 978-1-4799-1114-1. doi: 10.1109/IGARSS.2013.6723264.

Fabian. E. Bachl, Alex. Lenkoski, Thordis. L. Thorarinsdottir, and Christoph. S. Garbe. Bayesian motion estimation for dust aerosols. *ArXiv e-prints, arXiv:1308.0469, submitted to Annals of Applied Statistics*, August 2013b. under review.

L. Mark Berliner. Hierarchical Bayesian time series models. In Kenneth M. Hanson and Richard N. Silver, editors, *Maximum Entropy and Bayesian Methods*, volume 79 of *Fundamental Theories of Physics*, pages 15–22. Springer Netherlands, 1996. ISBN 978-94-010-6284-8. doi: 10.1007/978-94-011-5430-7_3.

Julian Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society Series B (Methodological)*, 36:192–236, 1974. ISSN 0035-9246. doi: 10.2307/2984812.

Julian Besag. On a system of two-dimensional recurrence equations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 43(3):302–309, 1981. ISSN 0035-9246.

Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006. ISBN 0387310738.

David Bolin. Spatial Matérn fields driven by non-Gaussian noise. *Scandinavian Journal of Statistics*, in press, 2013. ISSN 1467-9469. doi: 10.1111/sjos.12046.

David Bolin, Jonas Wallin, and Finn Lindgren. Multivariate latent Gaussian random field mixture models. *Preprint Department of Mathathematical Sciences, Chalmers University of Technology and Göteborg University*, 1, 2014.

Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23 (11):1222–1239, November 2001. ISSN 0162-8828. doi: 10.1109/34.969114.

Helen Brindley, Peter Knippertz, Claire Ryder, and Ian Ashpole. A critical evaluation of the ability of the spinning enhanced visible and infrared imager (SEVIRI) thermal infrared red-green-blue rendering to identify dust events: Theoretical analysis. *Journal of Geophysical Research: Atmospheres*, 117(D7):D07201, April 2012. ISSN 0148-0227. doi: 10.1029/2011JD017326.

T. Corpetti, E. Mémin, and P. Péréz. Dense estimation of fluid flows. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(3):365–380, March 2002. ISSN 0162-8828. doi: 10.1109/34.990137.

Noel A. C. Cressie and Christopher K. Wikle. *Statistics for Spatio-Temporal Data (Wiley Series in Probability and Statistics)*. Wiley, 1. edition, May 2011. ISBN 0471692743.

Yehia Eissa, Hosni Ghedira, Taha B. M. J. Ouarda, and Matteo Chiesa. Dust detection over bright surfaces using high-resolution visible SEVIRI images. In *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*, pages 3674–3677, July 2012.

Torbjørn Eltoft, Taesu Kim, and Te-Won Lee. On the multivariate Laplace distribution. *IEEE Signal Processing Letters*, 13(5):300–303, May 2006. ISSN 1070-9908. doi: 10.1109/LSP.2006.870353.

Ludwig Fahrmeir and Gerhard Tutz. *Multivariate Statistical Modelling Based on Generalized Linear Models*. Springer-Verlag, New York, second edition, April 2001. ISBN 0387951873.

J. Michael Fitzpatrick. The existence of geometrical density-image transformations corresponding to object motion. *Computer Vision, Graphics, and Image Processing*, 44(2):155–174, 1988. ISSN 0734-189X. doi: 10.1016/S0734-189X(88)80003-3.

Stuart Geman and Donald Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741, November 1984. ISSN 0162-8828. doi: 10.1109/TPAMI.1984.4767596.

Eric Gilleland, Johan Lindström, and Finn Lindgren. Analyzing the image warp forecast verification method on precipitation fields from the ICP. *Weather and Forecasting*, 25(4):1249–1262, March 2010. ISSN 0882-8156. doi: 10.1175/2010WAF2222365.1.

C. A. Glasbey and K. V. Mardia. A review of image-warping methods. *Journal of Applied Statistics*, 25(2):155–171, 1998. ISSN 0266-4763. doi: 10.1080/02664769823151.

T. Gneiting and A. E. Raftery. Weather forecasting with ensemble methods. *Science*, 310:248–249, October 2005. ISSN 0036-8075. doi: 10.1126/science.1115255.

Dale W. Griffin and Christina A. Kellogg. Dust storms and their impact on ocean and human health: Dust in Earth‘s atmosphere. *EcoHealth*, 1(3):284–295, 2004. ISSN 1612-9202. doi: 10.1007/s10393-004-0120-8.

Peter Guttorp and Tilmann Gneiting. Studies in the history of probability and statistics XLIX on the Matérn correlation family. *Biometrika*, 93(4):pp. 989–995, 2006. ISSN 0006-3444. doi: 10.1093/biomet/93.4.989.

Wolfgang Hackbusch. *Theorie und Numerik elliptischer Differentialgleichungen*. Teubner Studienbücher. Vieweg+Teubner Verlag, Leipzig, 2005. ISBN 3-519-02074-2. doi: 10.1007/978-3-322-99946-7.

B. Heinold, I. Tegen, K. Schepanski, and O. Hellmuth. Dust radiative feedback on Saharan boundary layer dynamics and dust mobilization. *Geophysical Research Letters*, 35(20):L20817, October 2008. ISSN 1944-8007. doi: 10.1029/2008GL035319.

Dominique Heitz, Etienne Mémin, and Christoph Schnörr. Variational fluid flow measurements from image sequences: synopsis and perspectives. *Experiments in Fluids*, 48(3):369–393, March 2010. ISSN 0723-4864. doi: 10.1007/s00348-009-0778-3.

Berthold K. P. Horn and Brian G. Schunck. Determining optical flow. *Artificial Intelligence*, 17(1–3):185–203, 1981. ISSN 0004-3702. doi: 10.1016/0004-3702(81)90024-2.

T. D. Jickells, Z. S. An, K. K. Andersen, A. R. Baker, G. Bergametti, N. Brooks, J. J. Cao, P. W. Boyd, R. A. Duce, K. A. Hunter, H. Kawahata, N. Kubilay, J. Laroche, P. S. Liss, N. Mahowald, J. M. Prospero, A. J. Ridgwell, I. Tegen, and R. Torres. Global iron connections between desert dust, ocean biogeochemistry, and climate. *Science*, 308:67–71, April 2005. ISSN 0036-8075. doi: 10.1126/science.1105959.

L. Klüser and K. Schepanski. Remote sensing of mineral dust over land with MSG infrared channels: A new bitemporal mineral dust index. *Remote Sensing of Environment*, 113(9):1853–1867, September 2009. ISSN 0034-4257. doi: 10.1016/j.rse.2009.04.012.

L. Knorr-Held and H. Rue. On block updating in Markov random field models for disease mapping. *Scandinavian Journal of Statistics*, 29(4):597–614, December 2002. ISSN 0303-6898. doi: 10.1111/1467-9469.00308.

Daphne Koller and Nir Friedman. *Probabilistic Graphical Models: Principles and Techniques - Adaptive Computation and Machine Learning*. The MIT Press, 2009. ISBN 9780262013192.

Kai Krajsek and Rudolf Mester. On the equivalence of variational and statistical differential motion estimation. In *IEEE Southwest Symposium on Image Analysis and Interpretation*, pages 11–15, 2006a. ISBN 1-4244-0069-4. doi: 10.1109/SSIAI.2006.1633712.

Kai Krajsek and Rudolf Mester. A maximum likelihood estimator for choosing the regularization parameters in global optical flow methods. In *IEEE International Conference on Image Processing*, pages 1081–1084, October 2006b. doi: 10.1109/ICIP.2006.312743.

Steffen L. Lauritzen. *Graphical Models (Oxford Statistical Science Series)*. Oxford University Press, USA, July 1996. ISBN 0198522193.

Itamar Lensky and Daniel Rosenfeld. Clouds-aerosols-precipitation satellite analysis tool (CAPSAT). *Atmospheric Chemistry and Physics*, 8(22):6739–6753, 2008. doi: 10.5194/acp-8-6739-2008.

Finn Lindgren, Håvard Rue, and Johan Lindström. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society Series B*, 73(4):423–498, 2011. ISSN 1467-9868. doi: 10.1111/j.1467-9868.2011.00777.x.

Chenyi Liu, Paul Fieguth, and Christoph S. Garbe. Background subtraction and dust storm detection. In *Geoscience and Remote Sensing Symposium, 2012. IGARSS 2012, Munich. IEEE International*, pages 2179–2181, 2012.

Bruce D. Lucas and Takeo Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of the 7th international joint conference*

on Artificial intelligence, volume 2, pages 674–679, San Francisco, CA, USA, April 1981. Morgan Kaufmann Publishers Inc.

Natalie M. Mahowald, Alex R. Baker, Gilles Bergametti, Nick Brooks, Robert A. Duce, Timothy D. Jickells, Nilgn Kubilay, Joseph M. Prospero, and Ina Tegen. Atmospheric global dust cycle and iron inputs to the ocean. *Global Biogeochemical Cycles*, 19(4):GB4025, December 2005. ISSN 1944-9224. doi: 10.1029/2004GB002402.

Sara Martino. *Approximate Bayesian Inference for Latent Gaussian Models*. PhD thesis, Norwegian University of Science and Technology, Department of Mathematical Sciences, Trondheim, 2007.

C. Marzban and S. Sandgathe. Optical flow for verification. *Weather and Forecasting*, 25(5):1479–1494, October 2010. ISSN 0882-8156. doi: 10.1175/2010WAF2222351.1.

J. A. Nelder and R. W. M. Wedderburn. Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)*, 135(3):370–384, 1972. ISSN 0035-9238. doi: 10.1007/978-1-4612-4380-9_38.

J. Pearl and A. Paz. Graphoids: A graph-based logic for reasoning about relevance relations. Technical Report R–53–L, Cognitive Systems Laboratory, University of California, Los Angeles, 1985.

Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1988. ISBN 0-934613-73-7.

P. Rivas-Perea, J. G. Rosiles, and M. Chacon. Traditional and neural probabilistic multispectral image processing for the dust aerosol detection problem. In *IEEE Southwest Symposium on Image Analysis Interpretation (SSIAI) 2010*, pages 169–172, May 2010. doi: 10.1109/SSIAI.2010.5483890.

Christian P. Robert. *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*. Springer-Verlag, 2nd edition, May 2007. ISBN 978-0-387-71598-8.

Christian P. Robert and George Casella. *Monte Carlo Statistical Methods*. Springer Texts in Statistics. Springer-Verlag, New York, second edition, 2004. ISBN 0-387-21239-6.

Daniel Rosenfeld, Yinon Rudich, and Ronen Lahav. Desert dust suppressing precipitation: A possible desertification feedback loop. *Proceedings of the National Academy of Sciences*, 98(11):5975–5980, May 2001. doi: 10.1073/pnas.101122798.

S. Roth and M. J. Black. On the spatial statistics of optical flow. In *Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV)*, volume 1, pages 42–49, October 2005. doi: 10.1109/ICCV.2005.180.

H. Rue and L. Held. *Gaussian Markov Random Fields: Theory and Applications*, volume 104 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London, 2005.

Håvard Rue and Sara Martino. Approximate Bayesian inference for hierarchical Gaussian Markov random fields models. *Journal of Statistical Planning and Inference*, 137(10):3177–3192, October 2007. ISSN 0378-3758. doi: 10.1016/j.jspi.2006.07.016.

Håvard Rue, Sara Martino, and Nicolas Chopin. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal Of The Royal Statistical Society Series B*, 71(2):319–392, April 2009. ISSN 1467-9868. doi: 10.1111/j.1467-9868.2008.00700.x.

Paul D. Sampson and Peter Guttorp. Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association*, 87 (417):108–119, March 1992. ISSN 0162-1459. doi: 10.2307/2290458.

R. Schefzik, T. L. Thorarinsdottir, and T. Gneiting. Uncertainty quantification in complex simulation models using ensemble copula coupling. *Statistical Science*, 28 (4):616–640, November 2013. ISSN 0883-4237. doi: 10.1214/13-STS443.

K. Schepanski. *Characterising Saharan Dust Sources and Export using Remote Sensing and Regional Modelling*. PhD thesis, Christian-Albrechts-Universit zu Kiel, 2009.

K. Schepanski, I. Tegen, B. Laurent, B. Heinold, and A. Macke. A new Saharan dust source activation frequency map derived from MSG-SEVIRI IR-channels. *Geophysical Research Letters*, 34(18):L13401, September 2007. ISSN 1944-8007. doi: 10.1029/2007GL030168.

K. Schepanski, I. Tegen, and A. Macke. Comparison of satellite based observations of Saharan dust source areas. *Remote Sensing of Environment*, 123(0):90–97, August 2012. ISSN 0034-4257. doi: 10.1016/j.rse.2012.03.019.

J. Schmetz, P. Pili, S. Tjemkes, D. Just, J. Kerkmann, S. Rota, and A. Ratier. An introduction to Meteosat Second Generation (MSG). *Bulletin of the American Meteorological Society*, 83(7):977–992, July 2002. ISSN 0003-0007. doi: 10.1175/1520-0477(2002)083⟨0977:AITMSG⟩2.3.CO;2.

Christoph Schnörr. Determining optical flow for irregular domains by minimizing quadratic functionals of a certain class. *International Journal of Computer Vision*, 6(1):25–38, April 1991. ISSN 0920-5691. doi: 10.1007/BF00127124.

Suzanne W. Seemann, Eva E. Borbas, Robert O. Knuteson, Gordon R. Stephenson, and Hung-Lung Huang. Development of a global infrared land surface emissivity database for application to clear sky sounding retrievals from multispectral satellite

radiance measurements. *Journal of Applied Meteorology and Climatology*, 47(1): 108–123, January 2008. ISSN 1558-8424. doi: 10.1175/2007JAMC1590.1.

E. P. Simoncelli, E. H. Adelson, and D. J. Heeger. Probability distributions of optical flow. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 310–315, June 1991. doi: 10.1109/CVPR.1991.139707.

Daniel Simpson, Finn Lindgren, and Hvard Rue. In order to make spatial statistics computationally feasible, we need to forget about the covariance function. *Environmetrics*, 23(1):65–74, 2012. ISSN 1099-095X. doi: 10.1002/env.1137.

Benjamin Sultan, Karima Labadi, Jean-Franois Gugan, and Serge Janicot. Climate drives the Meningitis epidemics onset in West Africa. *PLoS Medicine*, 2(1):E6, January 2005. ISSN 1549-1277. doi: 10.1371/journal.pmed.0020006.

Luke Tierney and Joseph B. Kadane. Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81(393): 82–86, March 1986. ISSN 01621459. doi: 10.2307/2287970.

P. Whittle. On stationary processes in the plane. *Biometrika*, 41(3/4):434–449, December 1954. ISSN 0006-3444. doi: 10.2307/2332724.

P. Whittle. Stochastic processes in several dimensions. *Bulletin of the International Statistical Institute*, 40:974–994, 1963.

Ke Xu, Christopher K. Wikle, and Neil I. Fox. A kernel-based spatio-temporal dynamical model for nowcasting weather radar reflectivities. *Journal of the American Statistical Association*, 100(472):1133–1144, 2005. ISSN 0162-1459. doi: 10.1198/016214505000000682.