The Annals of Applied Statistics
2015, Vol. 9, No. 3, 1298–1327
DOI: 10.1214/15-AOAS835

BAYESIAN MOTION ESTIMATION FOR DUST AEROSOLS[1]

BY FABIAN E. BACHL[*], ALEX LENKOSKI[†,2],
THORDIS L. THORARINSDOTTIR[†,2] AND CHRISTOPH S. GARBE[‡]

University of Bath[*], Norwegian Computing Center[†] and Heidelberg University[‡]

Dust storms in the earth's major desert regions significantly influence microphysical weather processes, the CO_2-cycle and the global climate in general. Recent increases in the spatio-temporal resolution of remote sensing instruments have created new opportunities to understand these phenomena. However, the scale of the data collected and the inherent stochasticity of the underlying process pose significant challenges, requiring a careful combination of image processing and statistical techniques. Using satellite imagery data, we develop a statistical model of atmospheric transport that relies on a latent Gaussian Markov random field (GMRF) for inference. In doing so, we make a link between the optical flow method of Horn and Schunck and the formulation of the transport process as a latent field in a generalized linear model. We critically extend this framework to satisfy the integrated continuity equation, thereby incorporating a flow field with nonzero divergence, and show that such an approach dramatically improves performance while remaining computationally feasible. Effects such as air compressibility and satellite column projection hence become intrinsic parts of this model. We conclude with a study of the dynamics of dust storms formed over Saharan Africa and show that our methodology is able to accurately and coherently track storm movement, a critical problem in this field.

1. Introduction. Dust storms are global meteorological phenomena originating from arid and semi-arid regions. They interfere with human modes of living and transportation, alter the radiation transmittance and circulation of the earth's atmosphere, and interact with microphysical cloud processes. Moreover, dust deposition provides vital nutrients for microorganisms that ultimately influence the CO_2-cycle. The detection of dust storms, the prediction of their development and the estimation of sources are therefore of immediate interest for a wide range of environmental applications. Remote sensing systems play an indispensable role in

Received August 2013; revised March 2015.

[1]Supported in part by the German Research Foundation (DFG) within the program "Spatio-Temporal Graphical Models and Applications in Image Analysis," Grant GRK 1653. The bulk of this research was conducted while the first author was a Ph.D. student in this program under the supervision of the fourth author.

[2]Supported in part by sfi[2], Statistics for Innovation, in Oslo.

Key words and phrases. Gaussian Markov random field, Horn and Schunck model, integrated continuity equation, integrated nested Laplace approximation (INLA), optical flow, remote sensing, satellite data, Saharan dust storm, storm tracking.

characterizing the dynamics of these systems, thereby providing the raw data that enables statistical analysis.

This article discusses the development of a comprehensive Bayesian hierarchical framework that uses remote sensing data to detect dust plumes, track their movement and pinpoint their source in a statistically sound manner. We show that a probabilistic approach is capable of coherently detecting the presence and absence of atmospheric aerosols. Furthermore, we show how standard models for aerosol flow can be linked to statistical models involving latent dependent random effects and then extend these basic models to incorporate more realistic features pertinent to dust storms. In particular, we develop a statistical method for estimating flow fields under flux that directly translates to the estimation of a GMRF in a hierarchical Bayesian model. While clear in retrospect, such a development is lacking in the applied community, with the effect that existing flow models fail to capture important features specific to aerosol transport. Our framework then allows a rich set of questions to be investigated both to pinpoint sources of dust storms and to track their motion.

The Meteosat series of satellites and, in particular, the Spinning Enhanced Visible and InfraRed Imager (SEVIRI) aboard the geostationary Meteosat-9 poses a unique opportunity, as it is the first time that the respective spatial and temporal coverage allows for the analysis of local and sub-daily processes of dust emission and transport. Alongside visible spectra, SEVIRI provides infrared measurements at frequencies from 3.9 to 13.4 μm every 15 minutes at a spatial resolution of 3 km at nadir. Figure 1 shows a visual depiction of the so-called SEVIRI falsecolor imagery (SFI), a common mode of visually assessing dust aerosols which forms the basis of our data.

Contemporary analysis of dust aerosols follows 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 SFI and aerosol optical depth (AOD). Here, the presence of dust is quantified by a combination of different SFI thresholds derived from case- and simulation-studies. In contrast, the work of Rivas-Perea, Rosiles and Chacon (2010) and Eissa et al. (2012) employs methods from machine learning and image processing by using neural nets to learn nonlinear dust detection criteria from a data set with labels set by a human expert.

From a statistical viewpoint, both approaches suffer shortcomings. Directly imposing thresholds partly based on expert opinion might lead to misleading conclusions due to human subjectivity. Also, neither Klüser and Schepanski (2009) nor Brindley et al. (2012) include quantification of uncertainty in their analysis. Neural nets, on the other hand, are directly driven by data and interpretable in a probabilistic sense. However, these methods are often criticized for a lack of transparency and nonphysical motivation, which in turn obfuscates scientific interpretability.

Further, none of the previously mentioned approaches imposes a coherent spatio-temporal structure. As a respective smoothness assumption can easily be

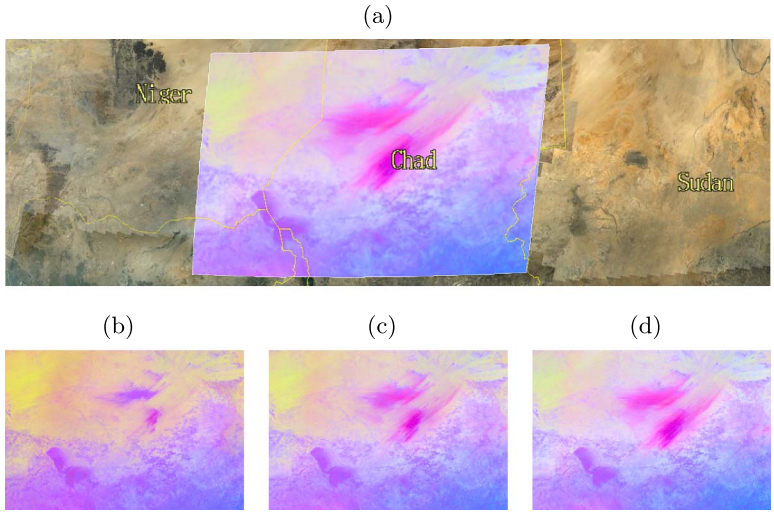

FIG. 1. (a) *SEVIRI falsecolor imagery according to Lensky and Rosenfeld* (2008) *superimposed on a Google Earth depiction of Niger, Chad and Sudan. The central pink area is a dust plume emerging on January* 18, 2010 *at about* 9.30 *am GMT over Chad. Panels* (b) *to* (d) *visualize the development of the plume at* 7.30 *am*, 8.30 *am and* 9.30 *am GMT, respectively.*

justified by the corresponding transport process, this omits valuable information. Previous attempts, for example, by Schepanski et al. (2007), to localize and characterize areas being sources of dust storms have to rely on human visual data inspection. Bachl and Garbe (2012) show that a data-driven estimation of the dust flux allows for automation of this process and can indicate dust source presence. Dust flux may also be employed to perform hazard forecasts, to interpolate areas with missing observation data (such as areas covered by clouds), or to validate atmospheric wind field-based models.

Various approaches in different scientific disciplines capture similar problems and are closely related to our methodology. Statistical approaches are predominantly driven by applications related to either the verification of numerical weather predictions or the issuing of so-called nowcasts, forecasts for very short lead-times; see, for example, Gilleland, Lindström and Lindgren (2010) and Xu, Wikle and Fox (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 prediction field both in terms of mislocalization 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, Wikle and Fox (2005) apply an 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 Optical Flow (OF) method of Horn and Schunck (1981), for brevity called HS-OF from here on. These methods have already entered the statistics community; see, for example, Marzban and Sandgathe (2010), who employ the connate OF approach of Lucas and Kanade (1981).

In this contribution we begin with the HS-OF method and illustrate how to formulate this approach as a Bayesian hierarchical model. This gives an interpretation of the flow as a latent Gaussian Markov random field with a precision hyperparameter of the imposed conditional autoregression model reflecting the intrinsic smoothness parameter of the HS-OF method. While the link is relatively straightforward, to the best of our knowledge, this is the first time that the full distributional aspects of the method and the associated uncertainty are taken into account. This perspective comes with several long- and short-term benefits. It allows for inference via computationally efficient integrated nested Laplace approximations (INLA) [Rue, Martino and Chopin (2009)] and leads to an extended interpretability of the flow field in terms of the physical nature of the phenomenon under consideration.

Our second contribution is to leverage the hierarchical Bayesian framework to overcome deficiencies in the HS-OF formulation. A typical quirk of statistical warping and optical flow is the underlying preservation assumption of the respective quantity along its trajectory. For remote assessment of dust aerosols (as well as other natural phenomena) this might lead to false conclusions, since gaseous solutions are compressible and only a nonbijective mapping of a three-dimensional quantity to a two-dimensional data space is at hand. As a remedy, we extend the HS-OF method to incorporate the water vapor related work of Corpetti, Memin and Perez (2002) and put the inherent integrated continuity equation (ICE) in a Bayesian hierarchical model context. As our work emphasizes by a simulation study, this considerably reduces errors in the estimated flow field. The main advantage of the ICE comes from the fact that it implicitly considers a multiplicative accumulation effect that is driven by the divergence of the flow field itself.

The article proceeds as follows. Section 2 offers a description of the data and background on the equipment used in their collection. Section 3 is twofold. As a first step it illustrates the basic thresholding concept for dust detection as well as our approach to employ a generalized linear model for this task. The Horn and Schunck method for motion estimation is then reviewed and extended by incorporating the integrated continuity equation, and a probabilistic interpretation of both approaches is provided. In Section 4 we evaluate our framework in three ways. First, we assess the detection method in comparison to thresholding and linear discriminant approaches. Section 4 then focuses on a simulation study analyzing Bayesian inference of the motion estimation techniques mentioned above. Finally, we show results of applying ICE motion estimation to dust detected from SEVIRI measurements, demonstrating forecasting capabilities and a method way to detect dust sources. Section 5 provides a discussion of our results and future work.

2. Data and operative products. 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). SEVIRI measures electromagnetic radiation at 12 visible and infrared spectra [Schmetz et al. (2002)]. Residing at 0 degrees of latitude, 0 degrees of longitude and a height of approximately 36 km, data are collected for up to approximately 80 degrees of deviation from nadir where the resolution is about 3×3 km. With the per-image scan time of 12 minutes and three minutes of calibration, this results in a 3712×3712 pixel per-channel imagery every 15 minutes.

The most dominant influence of dust aerosols on this data is to filter the infrared radiation, leaving the terrestrial surface in a frequency dependent fashion. This phenomenon is reflected by the channels $BT_{12.0}$, $BT_{10.8}$ and $BT_{8.7}$, where subscript denotes the respective frequency in microns. 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 results in popular operative products such as the SFI for which the red (R), green (G) and blue (B) visualization channels are defined as

$$\begin{aligned} R &= (\Delta T_{BR} + 4K)/(2K + 4K), \\ G &= (\Delta T_{BG}/15K)^{\gamma}, \\ B &= (BT_{10.8} - 261K)/(289K - 261K), \end{aligned}$$

where $\gamma = 0.4$ and K denotes the unit of brightness temperature in Kelvin. This maps the data to the interval $[0, 1]$ such that changes due to dust activity are most noticeable during on-screen inspection by experts [see Lensky and Rosenfeld (2008) for a detailed discussion on this and related visualization procedures]. Moreover, a study of Brindley et al. (2012) shows that this leads to a correlation between the tendency of the SFI to appear pink and the optical depth τ_{10} of the atmosphere at 10 μm being increased by the presence of dust aerosols. Recently, Ashpole and Washington (2012) proposed an extended thresholding scheme for dust detection given by

$$\Delta T_{BR} > 0K, \tag{1}$$

$$\Delta T_{BG} < 10K, \tag{2}$$

$$BT_{10.8} < 285K, \tag{3}$$

$$\Delta T_{BR} - M < -2K, \tag{4}$$

where M is a two-week cloud masked rolling mean of ΔT_{BR}. Alongside requiring the fixed conditions given in equations (1) and (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) removes artifacts coming from the latter. The last threshold is data dependent and serves two purposes. By requiring equation (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.

3. Methods. This section is twofold. We first discuss dust aerosol detection, that is, the task of assigning a given pixel of the SEVIRI imagery with a quantity representing the evidence for dust. The methodology we ultimately employ comes from a series of results [Bachl and Garbe (2012), Bachl, Fieguth and Garbe (2012, 2013)] in this area. However, for completeness we briefly review the main aspects of this literature below, which culminates in our final model depicted in (8).

Section 3.2 contains our main new methodological contribution. In this section we use the output of this dust detection model to infer the motion of the aerosol by modeling the underlying transport process relying on a differential perspective. We show that industry-standard methods, presently fit in a "engineering oriented" manner, have natural analogues in statistical models of dependent random effects. Further, these differential models can be extended nontrivially to incorporate more realistic features of aerosol transport without sacrificing computational feasibility in estimating their parameters.

One feature of the current methodology is that parameters of the two components of our model—dust detection and flow modeling—are fit separately. A procedure for estimating all parameters jointly is left for future work; see Section 5 for further discussion. All software related to this project is available from Bachl et al. (2015).

3.1. *Dust detection.* Let $\mathcal{S} \subset \mathbb{R}^2$ denote the image domain and assume we have a series of images obtained over the time interval $[0, T]$. Our first goal is to determine the dust indicator variable d_{xyt} with $d_{xyt} = 1$ if location $(x, y) \in \mathcal{S}$ is covered by a dust plume at time $t \in [0, T]$ and $d_{xyt} = 0$ otherwise. This assessment is made on the basis of the observation vector $\mathbf{I}_{xyt} = (I_{1xyt}, I_{2xyt}, I_{3xyt})$, where the three components of $\mathbf{I}_{xyt}$ correspond to the red, blue and green channels as discussed in Section 2. Since the surface in $\mathcal{S}$ is naturally varied, a critical component in determining d_{xyt} is the background appearance $\mathbf{A}_{xyt}$ at each location $(x, y) \in \mathcal{S}$ and time t when no dust or cloud cover exists. The background is compared to $\mathbf{I}_{xyt}$ to assess whether a dust plume covers the location at time t.

Our method of detecting dust aerosols is a progressive refinement of linear discriminant analysis (LDA). LDA infers projection coefficients r_i and an offset q such that the sign of

$$\eta(x, y, t) = q + \sum_{i=1}^{3} I_{ixyt} r_i \tag{5}$$

serves as a label for the dust content of a particular location. In Bachl and Garbe (2012), a three-level Bayesian hierarchical model is developed where the projection coefficients and intercepts are functions of the appearance estimate $\mathbf{A}_{xyt}$. In the first level, the dust indicator variable d_{xyt} is modeled via the logistic sigmoid

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

$$\eta(x, y, t) = \sum_{i=1}^{3} I_{ixyt} f_i^1(A_{ixyt}) + f_i^2(A_{ixyt}). \tag{6}$$

The second and third levels 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 semiparametrically by binning each component of $\mathbf{A}_{xyt}$ into 100 distinct bins taken over the range of each component over the image domain $\mathcal{S}$. These functionals are then modeled as second-order random walks (RW2s). That is,

$$f_i^1 \sim \mathcal{N}_{100}\big(0, Q_{\mathrm{RW2}}(\boldsymbol{\Theta}_i)\big),$$

$$f_i^2 \sim \mathcal{N}_{100}\big(0, Q_{\mathrm{RW2}}(\boldsymbol{\Xi}_i)\big),$$

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

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

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

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. A detailed discussion of this model is given in Section 3.4 of Rue and Held (2005).

Bachl, Fieguth and Garbe (2012) note that a drawback of this approach is that the signal noise in (6) is carried over in a linear fashion which can hamper consecutive motion estimation. As a remedy, they propose to shift the SFI to be a part of the domain of the latent functions such that

$$\eta(x, y, t) = \sum_{i=1}^{3} h_i(A_{ixyt}, A_{ixyt} - I_{ixyt}), \tag{7}$$

where the domain of $A_{ixyt} - I_{ixyt}$ is discretized in a manner similar to that of A_{ixyt}. The functions h_i are modeled as two-dimensional conditional autoregression (CAR) GMRFs [see Rue and Held (2005) for details] with

$$p(h_i) \propto \rho^{(n-1)/2} \exp\Bigg(-\frac{\rho}{2} \sum_{(l,m)\sim(j,k)} \big(h_i(l, m) - h_i(j, k)\big)^2\Bigg),$$

where "$\sim$" denotes the four nearest neighbors on the two-dimensional discretization grid of $A_{ixyt} \times (A_{ixyt} - I_{ixyt})$.

Yet, as discussed in Bachl, Fieguth and Garbe (2013), the estimation of the background radiation remains a critical aspect. Alongside the cyclic issue of requiring a criterion to mark a region as dust free, the radiative characteristics of this region generally vary even under pristine conditions. However, the vegetative properties of the largely unpopulated African continent significantly determines the general appearance of the SFI. The study therefore proposes to employ the monthly average surface emissivity $\mathbf{E}_{xyt}$ product at 8.4 μm according to Seemann et al. (2008). It strongly correlates with the vegetation and supersedes the anomaly indicating term $A - I$, hence,

$$\eta(x, y, t) = \sum_{i=1}^{3} g_i(I_{ixyt}, E_{ixyt}), \tag{8}$$

where the new functional g_i is modeled in a manner similar to h_i above. In the following, we refer to the model in (8) as the latent signal mapping (LSM) approach.

Our model therefore uses the functional (8) to predict the presence of dust. In practice, we are therefore required to determine several quantities, namely, the background appearance $\mathbf{A}_{xyt}$ or the emissivity $\mathbf{E}_{xyt}$, and subsequently fit a statistical model for $\eta(x, y, t)$. The Appendix gives the full model description and details regarding prior distributions. Estimation of this model is performed by using a large set of labeled training data; see Figure 2 for an example of one image used in our training set.

3.2. *Motion estimation*. 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, Mémin and Schnörr (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, for example, ozone and precipitation interpolation and forecasting. In particular, methods based on the perspective of warping have been in active development; see, for example, the review by Glasbey and Mardia (1998) and the work of Aberg et al. (2005).

However, to the best of our knowledge, the connection between probabilistic and variational approaches is reflected only by a few publications. Simoncelli, Adelson and Heeger (1991) point out the distributional aspects of the well-known Horn and Schunck (HS) method of optical flow [Horn and Schunck (1981)]. A maximum-a-posteriori approach to the free parameters of this method was illustrated by Krajsek and Mester (2006a) through the use of a Bayesian hierarchical model. Krajsek and Mester (2006b) 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.

Our approach utilizes the dust detection link function $\eta(x, y, t)$ estimated via the methods discussed in Section 3.1 as the primary input for modeling the flow field. Naturally, this quantity is estimated statistically and the most appropriate course of action would be to incorporate its uncertainty into the flow modeling. However, for computational reasons the posterior median of $\eta(x, y, t)$ is used instead. See the additional discussion of this feature in Section 5.

3.2.1. *The horn and schunck approach to optical flow.* Once the linear predictors of dust probability $\eta(x, y, t)$ are determined, it is helpful to model their dynamics in both space and time. This allows the projection of dust storm probabilities forward in time as well as "rewinding" the storm to determine its source.

As above, fix $(x, y) \in \mathcal{S}$. 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. As discussed in Section 1, we follow the motion estimation literature in our development and subsequently show that it is related to the Bayesian estimation of spatially dependent random effect models.

Like most motion estimation techniques, the HS-OF method is based on a preservation assumption concerning a photometric or geometric quantity in the image sequence. For a given triplet (x, y, t), suppose that $\eta(x, y, t) = k$. The HS-OF brightness constancy equation (BCE) then stipulates that there is a path in $\mathcal{S}$, $(x(r), y(r))$ for all $r \in [0, T]$ such that

$$\eta\big(x(r), y(r), r\big) = k. \tag{9}$$

Thus, the total derivative of the intensity function with respect to time vanishes. Assuming 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 holds that

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

where the dependence on (x, y, t) has been dropped.

This equation is under-determined, an issue known as the aperture problem. 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, nonsmoothness 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 a smoothness term integrated over the image domain $\mathcal{S}$. That is,

$$(u, v)(\alpha) = \operatorname*{argmin}_{u,v} L_{\mathrm{HS}}(\alpha),$$

where α is a regularization parameter and

$$L_{\mathrm{HS}}(\alpha) = \int_{\mathcal{S}} (\eta_t + u\eta_x + v\eta_y)^2 + \alpha^2\big(|\nabla u|^2 + |\nabla v|^2\big).$$

Existence and uniqueness of the minimizer were shown by Schnörr (1991) under mild restrictions on η and (u, v) in terms of Sobolev spaces.

In the discrete sense the BCE error term is 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,

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

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. The regularization term is discretized by approximating the integral over the image domain $\mathcal{S}$ with the Riemann sum over a regular grid $\mathcal{G} \subset \mathcal{S}$, that is,

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

in case of the horizontal flow gradient ∇u and equivalently for ∇v. Concomitantly, the partial derivatives are approximated by horizontal and vertical differences, that is, $u_x(i, j) = u_{ij} - u_{i-1,j}$ 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

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

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 intrinsic GMRF priors $p(u) = \mathcal{N}(\mathbf{0}, Q_u^{-1})$ and $p(v) = \mathcal{N}(\mathbf{0}, Q_v^{-1})$ for the latent flow fields if the precision matrices are defined via

$$Q_{ij}(\alpha) = \alpha^2 \begin{cases} n_i, & i = j, \\ -1, & i \sim j, \\ 0, & \text{otherwise}, \end{cases} \tag{10}$$

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

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

3.2.2. *The integrated continuity model.* While HS optical flow, and particularly the BCE assumption, is sufficient to model motion of rigid bodies in many areas of image processing, it is clearly insufficient in capturing the dynamics of $\eta(x, y, t)$. Constancy of image brightness implies that the flux of the quantity under consideration is divergence-free. This assumption is often violated for two reasons. On the one hand, the observed material itself might be compressible, as is the case for dust aerosols. Alternatively, even if incompressible fluids like water are considered, the imaging technique might deliver a two-dimensional projection of a three-dimensional process. Thus, even if this process obeys a divergence-free flow, the projection might miss strong sources and sinks due to the fluid convection through the layers of the z-axis.

The idea of the integrated continuity equation (ICE) [Corpetti, Memin and Perez (2002)] is that it relates the local intensity change to the flux of the quantity through the boundary surface of an infinitesimal volume:

$$\begin{aligned} 0 = \frac{\mathrm{d}}{\mathrm{d}t}\eta &= \eta_t + \operatorname{div}(\eta \mathbf{w}) \\ &= [\mathbf{w}, 1] \cdot \nabla \eta + \eta \operatorname{div}(\mathbf{w}). \end{aligned}$$

This equation also shows the connection to the BCE as it reduces to the former for incompressible materials when the divergence of $\mathbf{w}$ is zero and most importantly implies

$$\eta(x + u, y + v, t + \Delta t) = \eta(x, y, t) \exp\bigl(-\operatorname{div}(\mathbf{w})\bigr) \tag{11}$$

in a discrete setting where Δt is the time between two images. 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. Note that given a flow field $\mathbf{w}$, this equation is also easily employed to infer the temporal predecessor or successor of a given image, for example, by bilinear interpolation of the intensity values and scaling according to an approximation of the divergence.

Finally, we define the optical flow according to the ICE as the minimizer of the functional

$$L_{\mathrm{ICE}}(\alpha) = \int_{\mathcal{S}} \bigl([\mathbf{w}, 1] \cdot \nabla \eta + \eta \operatorname{div}(\mathbf{w})\bigr)^2 + \alpha^2 \bigl(|\nabla u|^2 + |\nabla v|^2\bigr).$$

Using the discrete divergence approximation

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

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

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

where again $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$. 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 from following 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.

It should be noted that both the HS and ICE methods follow an Eulerian perspective with an infinitesimal volume following the flow field. The principle difference is the surface of the volume. HS assumes that there is no flux through the surface, while the ICE approach imposes no such constraint and thus requires the additional modeling of the field's divergence. It is this additional feature that is able to appropriately capture the dynamics present in the flow fields.

In what follows we show that the 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, Martino and Chopin (2009)].

4. Applications. We now proceed with a series of studies that investigate the performance of the individual components of our framework and conclude with a set of case studies that show how the entire system performs at detecting and tracking dust storms. Section 4.1 focuses on the storm detection component—the model for determining $\eta(x, y, t)$—and compares our method with several reference methods. Section 4.2 then 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 4.3 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. As Section 4.5 shows, these forecasts can be improved upon by means of the Bayesian approach we take, that is, by postprocessing using marginal flow densities. Finally, we conclude in Section 4.6 with a procedure capable of tracing a dust storm back to its source and indicate the respective emission strength.

4.1. *Aerosol detection.* The basis of the following analysis is a SEVIRI data set spanning January 10–26, 2010, a period with several small- and large-scale dust events. By visual inspection we performed an extensive labeling of dusty and pristine regions. An example for a labeled frame of the sequence is given in Figure 2.

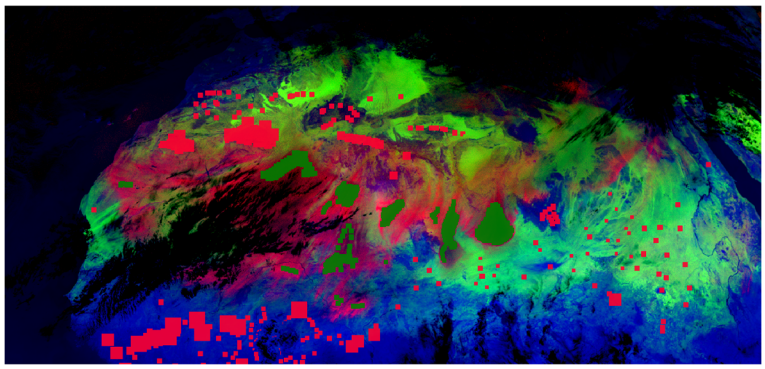

FIG. 2. *Training data in falsecolor representation. Red: pixels labeled as pristine. Green: pixels labeled as dusty.*

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 test set is employed to infer their dust probability and to compare it to the manually declared labels. These samples were flagged as dusty whenever the respective probability was above 0.5. This thresholding is performed in order to compare to the alternatives which are "pure classifiers," such as the ASH and ASH-no10.8 methods. 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 were computed. We compared the performance of four methods for estimating the probability of dust, the latent signal mapping (LSM) approach stated in equation (8), a simple linear discriminant analysis (LDA) and two thresholding approaches introduced by Ashpole and Washington (2012). In the 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 (1) and (2) hold. For the second method (ASH), also (3) is required to hold. Figure 3 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 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. Further, these effects vary considerably throughout the day, largely due to the manner that changes in overall illumination interact with the fixed boundaries of these

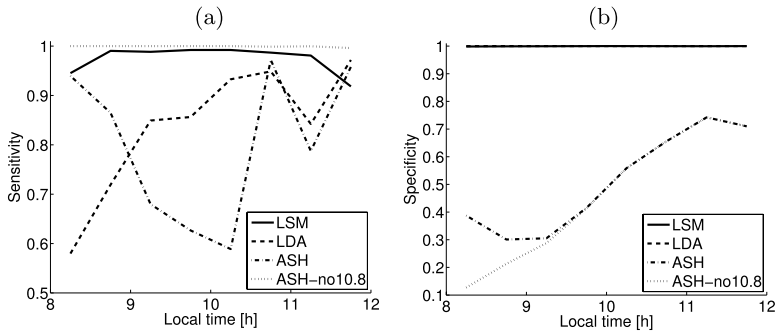

FIG. 3. *Cross-validation results for pixel-wise dust detection under the LSM emissivity approach, linear discriminant analysis and the two thresholding methods of Ashpole and Washington (2012). The plots show the percentage of correctly classified* (a) *dusty pixels and* (b) *pristine pixels, stratified by the hour of the day.*

methods. 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, Fieguth and Garbe (2013) and justify our use of the LSM emissivity modeling approach in (8) on these data.

Figure 4 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 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 cutoffs 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.

The data displayed in Figure 4 also demonstrate 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 are displayed in the bottom row of Figure 4. 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 8.3 μm radiation. In

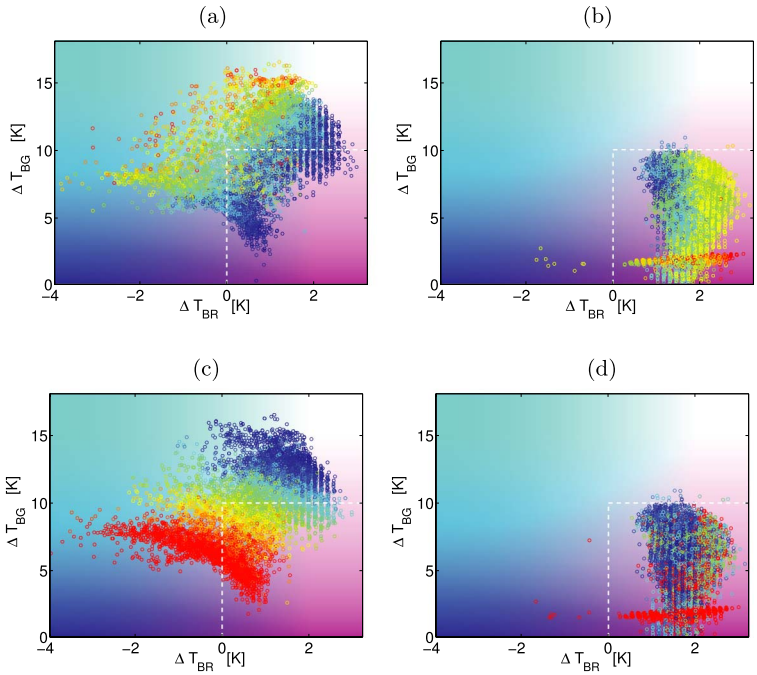

FIG. 4. *Green channel intensity* (*x-axis*) *versus red channel intensity* (*y-axis*) *of the labeled training data. Left column*: *pixels labeled as pristine*; *right column*: *pixels labeled as dusty*; *top row*: *points are colored by local time of the day, from blue* (*early*) *to red* (*later*); *bottom row*: *points are colored by emissivity, from blue* (*low*) *to red* (*high*). *The white dashed lines indicate the "no*10.8*" thresholding of Ashpole and Washington* (2012) *and the background coloring shows the appearance of a pixel assuming a fully saturated blue channel. As the entire data set is very large, each plot shows a random subsample of the full data set.*

combining this information with channel intensity in the LSM approach, we thus achieve an improved classification in the early-morning data.

4.2. *Simulation study*: *Aerosol flow.* 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 illustrate the difference between the two approaches. Figure 5 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 6 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 6 shows several interesting

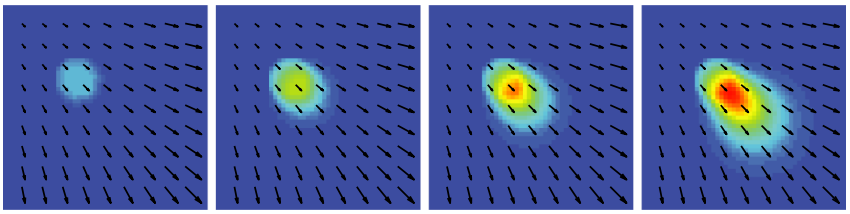

FIG. 5. *Synthetic image sequence of a dust plume and aerosol flow.*

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. Remember that the posterior flow field under the Bayesian approach is an average with weights according to the posterior of the tuning parameter. When the BCE is violated, this posterior may yield little information (e.g., it remains flat) or put an unreasonable amount of mass extremely close to zero. This way, severe degeneracies of the flow field can occur. For instance, we might obtain a field that is constant but points in the wrong direction.

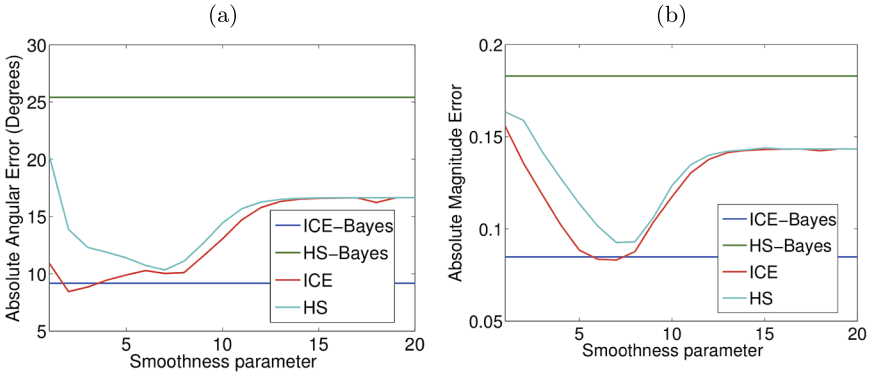

FIG. 6. *Quantification of* (a) *absolute angular and* (b) *absolute magnitude error of aerosol flow estimation for the synthetic image sequence in Figure 5. The plots compare the errors of the ICE and the HS methods under both standard and Bayesian inferences as a function of the smoothness parameter* α.

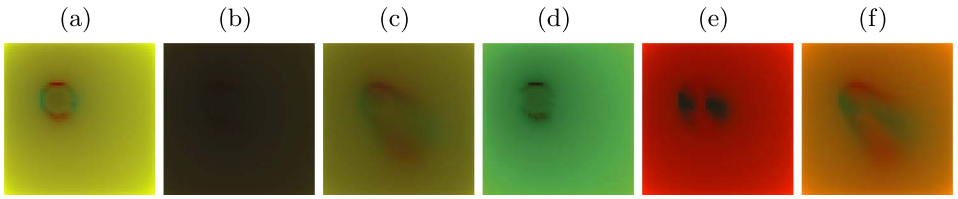

FIG. 7. *Marginal variances of the flow field vectors derived via the Bayesian ICE* [(a) *to* (c)] *and HS* [(d) *to* (f)] *approaches applied to the simulated data frames shown in Figure* 5. *Red coloring represents the uncertainty in the horizontal flow field component, while green expresses the same for the vertical component. Regions with low uncertainty in both components appear as black* [*e.g., panel* (b)], *while equally high variance is shown in yellow.*

These outliers overwhelm the sensible points of the posterior distribution, thereby leading to the poor performance observed.

A clear motivation to apply Bayesian hierarchical modeling and inference with INLA is the straightforward assessment of the marginal distributions of the latent fields. Figure 7 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 applied to the simulated data shown in Figure 5.

From Figure 7(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 due to the aperture problem mentioned in Section 3.2.1. Information about the direction of the motion can only be obtained in the direction of the gradient of the image sequence. Motion perpendicular to the gradient is locally not accessible and can thus not be detected by the model. Hence, as the vertical gradient is prominent at the lower and upper boundary of the dust source, the respective motion variance is low (less green color content), while the horizontal motion component is largely unknown and leads to a red coloring. Similarly, the left and right boundary of the plume show a high variance in the vertical direction (green coloring), while the horizontal motion is accessed with comparably large certainty.

The outer region as well as inside of the dust source behaves as expected as well. In both no gradient is present, which leads to large uncertainty in both components. As a mixture of green and red, these regions therefore appear yellow in the figure. This phenomenon is most prominent 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 Figure 7(d) to (f), the marginal variances can be very informative in terms of misspecifications of the model as well. In Figure 7(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 7(f) the situation is more balanced, but still predominance of the uncertainty in the horizontal (red) component can be observed. Last, the central area of the dust source region in this figure seems slightly more pronounced than in Figure 7(c). This is an additional indicator for the fact that, in particular, the HS approach struggles to reflect dust source effect, for example, the influx of dust mass into the atmosphere.

4.3. *Case studies*: *Detection and flow estimation*. 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 framework during the evolution of two separate dust storms. Figure 8 shows dust storms that occurred during January 8, 2010 and January 16, 2010. 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 Figure 8(a) and (b). 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. Second, we see why the ICE method is preferred over standard HS. There is considerably more regularity to the estimated flow field in the respective third rows than the second rows, 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 8(a). Given the results in Figure 8, we proceed in the rest of our studies by only considering the Bayesian ICE model.

4.4. *Forecasting dust event evolution*. 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 9, 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 (11) implies a straightforward 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.

First, we compute the dust predictor and flow with respect to the imagery of 11.45 h and 12.00 h GMT. Figure 9(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 northeastern Niger predominantly moving to the south, (B) one over southern Niger moving in a southern and western direction and, last, (C) a plume emerging at the borders between Algeria, Niger and Mali moving westward in the direction of central Mali. Then we extrapolate the field through iterative application of equation (11) 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

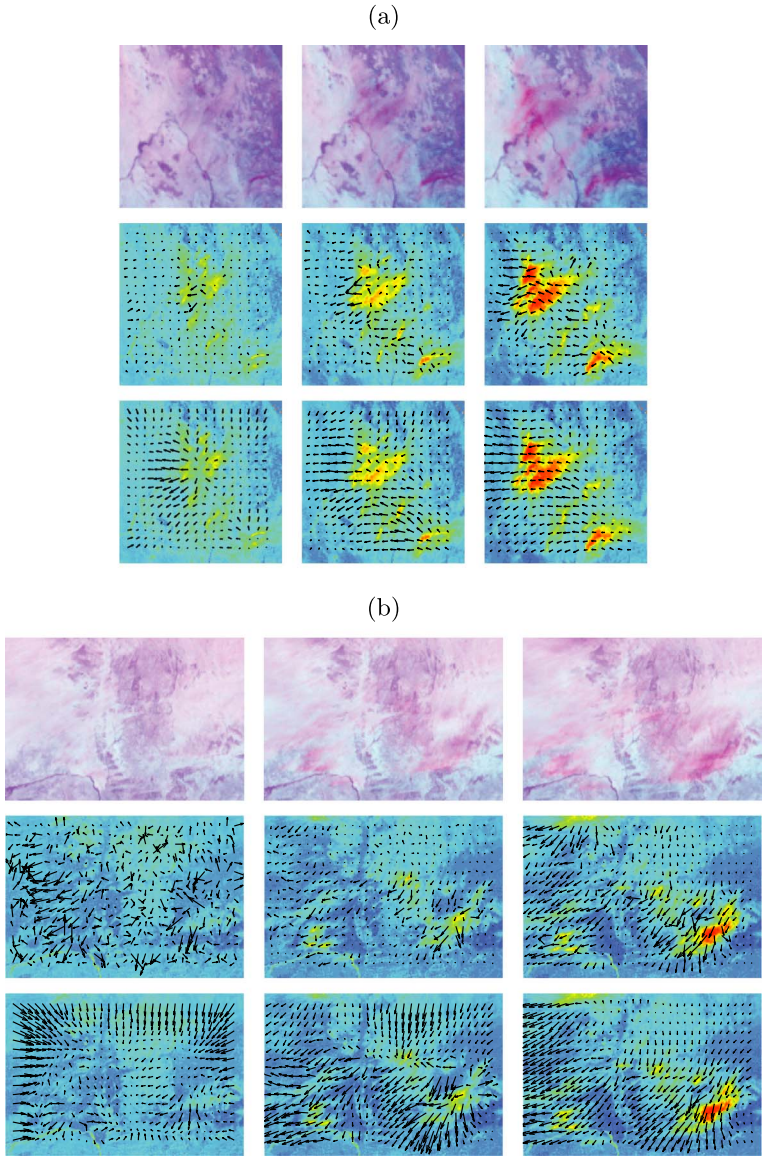

FIG. 8. *Dust plumes on* (a) *January* 8, 2010 *at* 7.15 *am*, 8.30 *am and* 11 *am GMT*; (b) *January* 16, 2010 *at* 10.15 *am*, 11.45 *am and* 1 *pm GMT. Top rows*: *observed satellite data in false color*; *middle rows*: *pixel-wise LSM probability of dust estimates overlaid with the Bayesian HS flow field*; *bottom rows*: *same pixel-wise probability of dust estimates as above now overlaid with the Bayesian ICE flow field.*

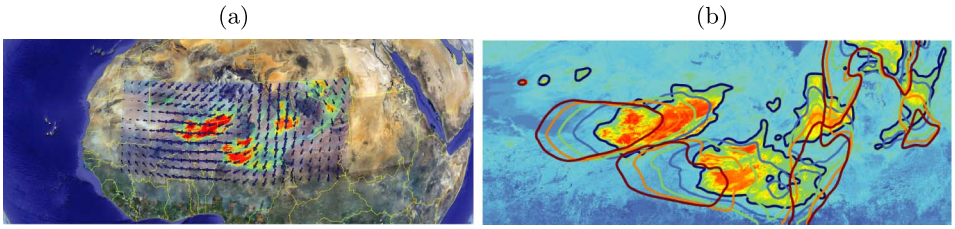

FIG. 9. *Forecast of a dust event I. Panel* (a) *shows the linear predictor of a dust event over northern Africa on January* 17, 2010, 12 *h GMT as a transparent overlay on Google Earth imagery including country borders and the flow field computed with the ICE method. Panel* (b) *depicts the contour 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.*

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 9(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 toward 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 10 shows how our forecast gains accuracy with a decreasing prediction interval. The actual dust density at January 18, 2010 12.00 h GMT is shown in

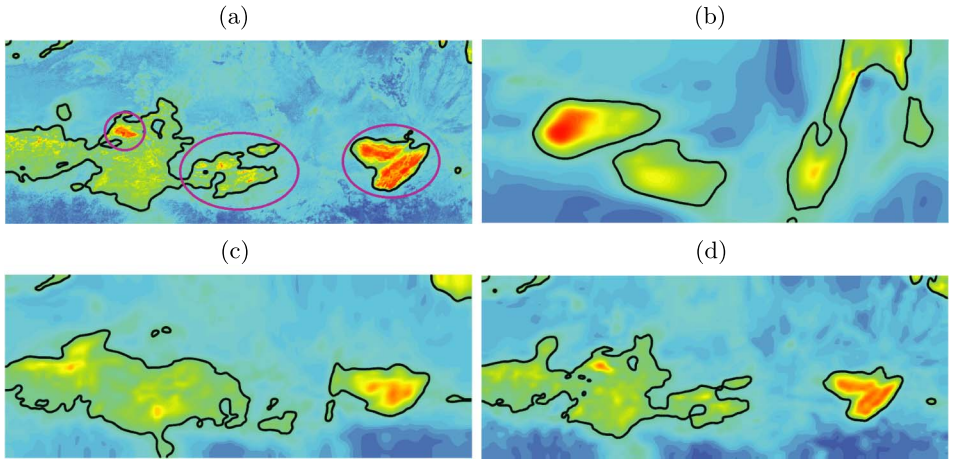

FIG. 10. *Forecasted dust densities. Panel* (a) *shows the actual predictor for dust density on January* 18, 12 *h GMT with a* 30% *contour line* (*black*). *The pink ellipses mark dust plumes emerging during the morning of this day. Panels* (b), (c) *and* (d) *show the dust density forecasted for this time based on data from* 24, 3 *and* 1 *hour before*, *respectively.*

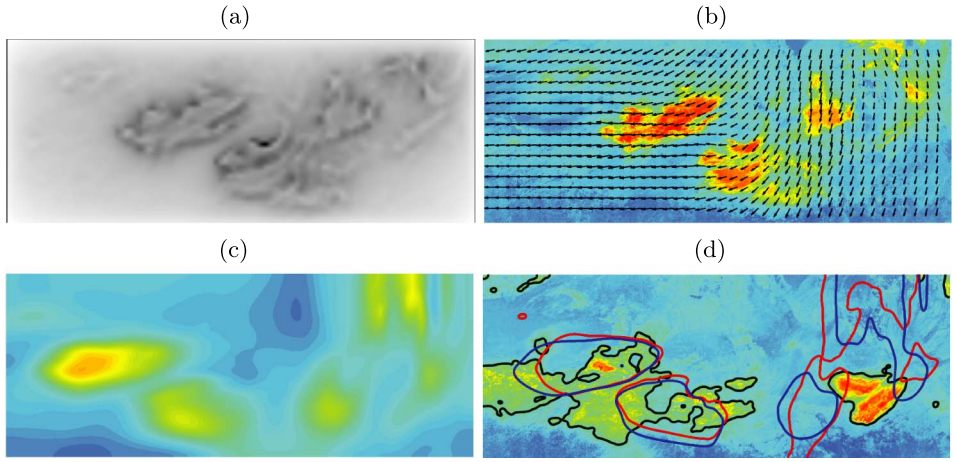

FIG. 11. *Postprocessing of flow estimates. Panel* (a) *shows the sum of the marginal flow component precisions derived for the dust forecast procedure elaborated in Section* 4.4. *These are employed to weight the local linear fit term of the ICE method, which leads to the flow field shown in panel* (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.*

Figure 10(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 24 h forecast depicted in Figure 10(b), from which another observation can be made. The forecasts of plumes (B) and (C) do not appear as widespread 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 3 h and 1 h forecasts shown in Figure 10(c) and (d) gain accuracy. Both indicate the new dust event in the southeast and the 1 h forecast also picks up the two weaker events.

4.5. *Postprocessing using marginal densities.* 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 noninformative regions. Consider again the dust event of January 17, 2010 and Figure 11(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:

$$L_{\mathrm{ICE^W}}(\alpha) = \int_{\mathcal{S}} \big(q(\gamma_1 + \gamma_2)\big)^2 \big([\mathbf{w}, 1] \cdot \nabla\eta + \eta \operatorname{div}(\mathbf{w})\big)^2 + \alpha^2\big(|\nabla u|^2 + |\nabla v|^2\big),$$

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 11(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 southeastern parts of the area under investigation, and also in between dust plumes. Figure 11(c) and (d) show visibly that this aides the forecast process and mitigates aforementioned problematic effects. In particular, plumes (B) and (C) move faster toward 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 southeastern 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.

4.6. *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.

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, Tegen and Macke (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, for example, studies that rely on

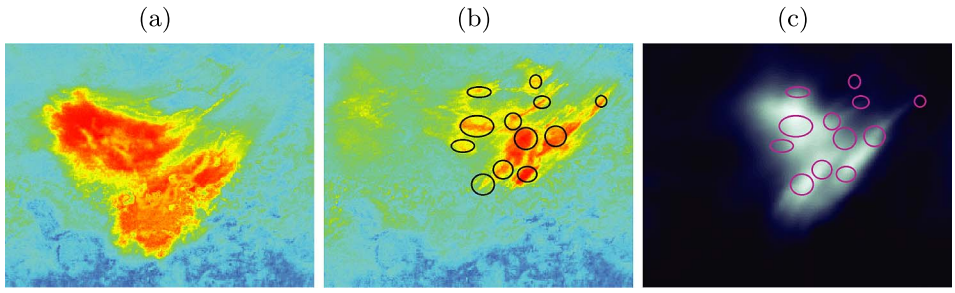

FIG. 12. *Spatial estimation of dust emission strength. Panel* (a) *shows the linear predictor of a dust even on January* 18, 2010 *at* 15.00 *h GMT over the Bodédé depression in northern Africa. This dust plume originates from a cluster of source regions identified by visual inspection of the image 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* (c) *to the presumed origin shown in panel* (d) *and serves as a spatial estimator for the dust emission strength.*

wind field averages and Lagrangian trajectories to trace back dust to its origin [Alonso-Pérez et al. (2012)].

Figure 12 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 signs of the event are visible in the data at 6.15 h GMT and the plume reaches its maximal extent at around 15.00 h GMT. We compute the flow of the plume for the whole period and then use equation (11) to transport the predictor of the imagery at 15.00 h GMT [see Figure 12(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 12(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 8.15 h 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 backward according to the determined flow field leads to the emission strength estimate depicted in Figure 12(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 northeastern 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 comparably small, in particular, in the direction orthogonal to the wind field that drives them. The imagery gradient necessary for our flow estima-

tion 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.

5. Discussion. We have outlined a Bayesian framework for detecting and tracking dust storms that significantly expands the existing methodology used in the remote sensing and earth observation communities for addressing such problems. The approach makes several developments, including a superior dust detection methodology, a link between the classical literature of optical flow and GMRFs—which incidentally shows how Bayesian estimation can alleviate issues related to the setting of tuning parameters—and the use of the ICE to model flow fields where an assumption of brightness constancy is inconsistent with the physical process. Simulation studies have shown the improved performance of both our storm detection framework and the Bayesian ICE model over existing procedures and real-world examples have shown the implications of this improvement. Furthermore, the use of the Bayesian approaches offers an automatic way of tuning smoothing parameters that appear to achieve nearly optimal levels of smoothing without the need for extensive cross-validation.

Considerable work remains, both from the application and methodological perspectives. The model for η appears to work quite well in our current data, but it could be extended in several obvious manners. The most useful of these would be to make the estimates of η not just depend 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 effort to such coupled estimation proved challenging—however, as the computational methodology for estimating such models 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, the former might be a critical feature to prevent a degradation of detection performance. Finally, the current two-dimensional model clearly misses the three-dimensional nature of dust storms. This reduction is performed since our data are column data, however, a latent understanding of the height of the dust could extend the model's capabilities.

The model relies on two main components whose parameters are currently estimated separately. Namely, a dust detection model is first trained and the fitted values of this model are then fed into a model for flow. A major next step would be to jointly estimate the detection and flow models, thereby feeding uncertainty in the detection into the flow estimation. In early stages of this project, we experimented with such a joint approach. However, we found the computational burden from such an approach to substantially outweigh the modest—at best—improvement in flow estimation that resulted. As new data and modeling scenarios are entertained, it is possible that a joint estimation strategy can yield greater improvement and should therefore be considered.

Regarding our procedure for estimating flow fields, a general smoothness assumption or even local constancy as within the Lucas and Kanade (1981) approach can in most cases be justified. Here, it is particularly appealing that the work of Lindgren, Rue and Lindström (2011) as well as Simpson, Lindgren and Rue (2012) reveals links between Gaussian fields and GMRFs via stochastic partial differential equations (SPDEs). Future approaches may find this link as a mode to refine the prior of the GMRF in terms of expressing a transport phenomenon via its SPDE, and thereby gain further insight into how it is reflected by the given data.

The connection to continuously modeled phenomena also comes up at the methodological intersection with image-processing methods. Traditionally, inferring the HS optical flow was subject to solving a variational formulation of the problem via the corresponding Euler–Lagrange equations. Most importantly, the variational perspective leads to further insight about the properness of the resulting GMRF with respect to the function space the data are sampled from. As shown by Schnörr (1991), relatively mild conditions, namely, a mildly restricted Sobolev space, are sufficient to guarantee this properness. It should also be mentioned that the likelihood term and respective choices of the error penalty of the HS optical flow and related methods has consistently been subject to several studies. Here, the corresponding flexibility of the GLM formulation and the INLA methodology might excel in further in-depth analyses.

While showing that remote sensing equipment can be used to detect and track dust storms was our initial goal, there are 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, Thorarinsdottir and Gneiting (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.

APPENDIX: FULL MODEL DESCRIPTIONS

A.1. The dust detection model. This section explicitly discusses the statistical formulation and estimation of the dust detection model $\eta(x, y, t)$. The model takes two inputs I_{ixyt} and E_{ixyt}, both of which are integral and indicate which bin the associated values are placed in, out of 100 potential bins. Thus, $I_{ixyt}, E_{ixyt} \in \{1, \ldots, 100\} = \mathbb{X}$. The dust link function is then modeled by

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

where

$$g_i : \mathbb{X} \times \mathbb{X} \to \mathbb{R}.$$

Each g_i can be represented by a 100×100 matrix $\mathbf{G}_i$ where $(\mathbf{G}_i)_{kl} = g_i(k, l)$. Each functional g_i is the modeled semiparametrically via

$$\text{vec}\{\mathbf{G}_i\} \sim \mathcal{N}_{100^2}(\mathbf{0}, \mathbf{\Theta}_\rho^{-1}),$$

where $\mathbf{\Theta}_\rho$ is a $100^2 \times 100^2$ sparse precision matrix with zeroes according a graph $\mathcal{G} = (V, E)$. The graph $\mathcal{G}$ is such that $V = (1, \ldots, 100^2)$ and each $v \in V$ can be mapped to a pair $(l, m) \in \mathbb{X} \times \mathbb{X}$. The edge set of $\mathcal{G}$ is such that $(v, v') \in E$ only when the corresponding pairs (l, m) and (l', m') have $l = l'$ and/or $m = m'$. Let $nb(v)$ be the number of edges in E involving the vertex $v \in V$. Given this construction, the elements of Θ_ρ are such that

$$(\mathbf{\Theta}_\rho)_{vv} = \rho nb(v),$$

while for $v \neq v'$ with $(v, v') \in \mathcal{G}$,

$$(\mathbf{\Theta}_\rho)_{vv'} = -\rho$$

with $(\mathbf{\Theta}_\rho)_{v,v'} = 0$ when $(v, v') \notin \mathcal{G}$. The elements of each $\mathbf{G}_i$ are modeled independently of the others. Thus, the full model is

$$\begin{aligned} pr(d_{xyt} = 1) &= \frac{\exp(\eta(x, y, t))}{1 + \exp(\eta(x, y, t))}, \\ \eta(x, y, t) &= \sum_{i=1}^{3} g_i(I_{ixyt}, E_{ixyt}), \\ \text{vec}(\mathbf{G}_i) &\sim \mathcal{N}_{100^2}(0, \Theta_{\rho_i}^{-1}), \qquad i \in \{1, \ldots, 3\}, \\ \rho_i &\sim \text{LogGamma}(1, 0.1), \qquad i \in \{1, \ldots, 3\}. \end{aligned}$$

A.2. The flow model. Recall at this junction that the linear component from the dust detection model $\eta(x, y, t)$ is now taken as given for all locations in $\mathcal{S}$ and time points. Since the majority of points in $\mathcal{S}$ have a vanishingly small value of $\eta(x, y, t)$, we find a rectangular subdomain which contains all points with a nonnegligible η value. That is, we find $\mathcal{S}_0 \subset \mathcal{S}$ such that $(x, y) \in \mathcal{S}_0$ implies $x_{\min} < x < x_{\max}$, $y_{\min} < y < y_{\max}$ for some $x_{\min}, x_{\max}, y_{\min}, y_{\max}$ chosen such that $(x, y) \notin \mathcal{S}_0$ implies that $\eta(x, y, t) < a$ for all t. In practice, we set $a = -5$, but results are insensitive to reasonable choices of this cutoff.

Once the subdomain of interest $\mathcal{S}_0$ is formed, the dust detection link $\eta(x, y, t)$ must be scaled down so that flow in a given time point cannot span more than one pixel as discussed in Section 3.2. In our applications a three-pixel downscaling typically proved sufficient. This means that we form a subgrid $\mathcal{S}_0^1 \subset \mathcal{S}_0$ such that $(x, y), (w, z) \in \mathcal{S}_0^1$ only when $|x - w| > 1$ and $|y - z| > 1$ and for each $(w, z) \in \mathcal{S}_0$ there exists a point $(x, y) \in \mathcal{S}_0^1$ such that $|x - w| \leq 1$ and $|y - z| \leq 1$. Relative to $\mathcal{S}_0^1$, we form the downscaled dust predictor

$$\tilde{\eta}(x, y, t) = \sum_{l \in \{-1,0,1\}} \sum_{m \in \{-1,0,1\}} \eta(x + l, y + m, t), \qquad (x, y) \in \mathcal{S}_0^1. \tag{12}$$

This way, $\tilde{\eta}$ has a third of the spatial resolution of η. In the case of the simulated data, no averaging was performed, as the data was generated in a way that the artificial dust plume does not move more than one pixel per time step.

The next postprocessing step uses $\tilde{\eta}$ to form the numerical derivatives. The spatial gradient was computed by the temporal average of the forward differences:

$$\begin{aligned}\tilde{\eta}_x(x,y,t) = \tfrac{1}{2}\{&\big(\tilde{\eta}(x+1,y,t) - \eta(x,y,t)\big) \\ &+ \big(\eta(x+1,y,t+1) - \eta(x,y,t+1)_{x,i,j,t+1}\big)\},\end{aligned} \tag{13}$$

$$\begin{aligned}\tilde{\eta}_y(x,y,t) = \tfrac{1}{2}\{&\big(\tilde{\eta}(x,y+1,t) - \tilde{\eta}(x,y,t)\big) \\ &+ \big(\tilde{\eta}(x,y+1,t+1) - \tilde{\eta}(x,y,t+1)\big)\}.\end{aligned} \tag{14}$$

Similarly, the temporal partial derivative is simply

$$\tilde{\eta}_t(x,y,t) = \tilde{\eta}(x,y,t+1) - \tilde{\eta}(x,y,t).$$

These derivatives are then used as data to estimate the downscaled vector fields $\tilde{u}(x,y,t)$ and $\tilde{v}(x,y,t)$ for $(x,y) \in \mathcal{S}_0^1$. This is modeled as

$$-\tilde{\eta}_t(x,y,t) = \tilde{\eta}_x(x,y,t)\tilde{u}(x,y,t) + \tilde{\eta}_y(x,y,t)\tilde{v}(x,y,t) + \nu(x,y,t)$$

separately for each time point t, where $\nu(x,y,t) \sim N(0, 1\text{e}{-4})$ is an i.i.d. white noise process reflecting instrument error. Hence, $\tilde{u}(x,y,t)$ and $\tilde{v}(x,y,t)$ are now random parameters to be estimated and each can be considered a matrix $\mathbf{U}_t$ and $\mathbf{V}_t$ of size $(x_{\max} - x_{\min}) \times (y_{\max} - y_{\min})$. Following similar steps as in the section above, we then place the prior

$$\begin{aligned}\operatorname{vec}(\mathbf{U}_t) &\sim \mathcal{N}\big(\mathbf{0}, \mathbf{Q}_{\alpha_u}^{-1}\big), \\ \operatorname{vec}(\mathbf{V}_t) &\sim \mathcal{N}\big(\mathbf{0}, \mathbf{Q}_{\alpha_v}^{-1}\big),\end{aligned}$$

where $\mathbf{Q}_{\alpha_u}, \mathbf{Q}_{\alpha_v}$ have the structure given in (10) and $\alpha_u, \alpha_v \sim \text{LogGamma}(1, 0.1)$ in the prior. Once the posterior distribution of $\tilde{u}$ and $\tilde{v}$ are determined over $\mathcal{S}_0^1$, the values for $u(x,y,t)$ on $\mathcal{S}_0$ are found via linear interpolation.

Acknowledgments. We thank the Editor, Associate Editor and three anonymous reviewers for their helpful comments.

SUPPLEMENTARY MATERIAL

Software (DOI: 10.1214/15-AOAS835SUPP; .zip). All software related to this project is available as supplemental material provided in Bachl et al. (2015). For an up-to-date version check the corresponding author's website, www.nr.no/~lenkoski.

REFERENCES

ABERG, S., LINDGREN, F., MALMBERG, A., HOLST, J. and HOLST, U. (2005). An image warping approach to spatio-temporal modelling. *Environmetrics* **16** 833–848. MR2216654

ALONSO-PÉREZ, S., CUEVAS, E., QUEROL, X., GUERRA, J. C. and PÉREZ, C. (2012). African dust source regions for observed dust outbreaks over the subtropical eastern North atlantic region, above 25B0N. *Journal of Arid Environments* **78** 100–109.

ASHPOLE, I. and WASHINGTON, R. (2012). An automated dust detection using SEVIRI: A multiyear climatology of summertime dustiness in the central and western Sahara. *Journal of Geophysical Research*: *Atmospheres* **117** D08202.

BACHL, F. E., FIEGUTH, P. and GARBE, C. S. (2012). A Bayesian approach to spaceborn hyperspectral optical flow estimation on dust aerosols. In *Proceedings of the International Geoscience and Remote Sensing Symposium* 2012 256–259. IEEE, New York.

BACHL, F. E., FIEGUTH, P. and GARBE, C. S. (2013). Bayesian inference on integrated continuity fluid flows and their application to dust aerosols. In *Proceedings of the International Geoscience and Remote Sensing Symposium* 2013 2246–2249. IEEE, New York.

BACHL, F. E. and GARBE, C. S. (2012). 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"* (L. Ouwehand, ed.). *ESA Special Publication* **703** S1–S3. European Space Agency. European Space Agency Communications, Frascati, Italy.

BACHL, F. E., LENKOSKI, A., THORARINSDOTTIR, T. L. and GARBE, C. S. (2015). Supplement to "Bayesian motion estimation for dust aerosols." DOI:10.1214/15-AOAS835SUPP.

BESAG, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **36** 192–236. MR0373208

BRINDLEY, H., KNIPPERTZ, P., RYDER, C. and ASHPOLE, I. (2012). 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** D07201.

CORPETTI, T., MEMIN, E. and PEREZ, P. (2002). Dense estimation of fluid flows. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24** 365–380.

EISSA, Y., GHEDIRA, H., OUARDA, T. B. M. J. and CHIESA, M. (2012). Dust detection over bright surfaces using high-resolution visible SEVIRI images. In *Proceedings of the International Geoscience and Remote Sensing Symposium* 2012 3674–3677. IEEE, New York.

GILLELAND, E., LINDSTRÖM, J. and LINDGREN, F. (2010). Analyzing the image warp forecast verification method on precipitation fields from the ICP. *Weather and Forecasting* **25** 1249–1262.

GLASBEY, C. A. and MARDIA, K. V. (1998). A review of image-warping methods. *J. Appl. Stat.* **25** 155–171.

GNEITING, T. and RAFTERY, A. E. (2005). Atmospheric science. Weather forecasting with ensemble methods. *Science* **310** 248–249.

HEITZ, D., MÉMIN, E. and SCHNÖRR, C. (2010). Variational fluid flow measurements from image sequences: Synopsis and perspectives. *Experiments in Fluids* **48** 369–393.

HORN, B. K. P. and SCHUNCK, B. G. (1981). Determining optical flow. *Artificial Intelligence* **17** 185–203.

JICKELLS, T. D., AN, Z. S., ANDERSEN, K. K., BAKER, A. R., BERGAMETTI, G., BROOKS, N., CAO, J. J., BOYD, P. W., DUCE, R. A., HUNTER, K. A., KAWAHATA, H., KUBILAY, N., LAROCHE, J., LISS, P. S., MAHOWALD, N., PROSPERO, J. M., RIDGWELL, A. J., TEGEN, I. and TORRES, R. (2005). Global iron connections between desert dust, ocean biogeochemistry, and climate. *Science* **308** 67–71.

KLÜSER, L. and SCHEPANSKI, K. (2009). Remote sensing of mineral dust over land with MSG infrared channels: A new bitemporal mineral dust index. *Remote Sensing of Environment* **113** 1853–1867.

KRAJSEK, K. and MESTER, R. (2006a). A maximum likelihood estimator for choosing the regularization parameters in global optical flow methods. In *IEEE International Conference on Image Processing* 1081–1084. IEEE, New York.

KRAJSEK, K. and MESTER, R. (2006b). On the equivalence of variational and statistical differential motion estimation. In *IEEE Southwest Symposium on Image Analysis and Interpretation* 11–15. Denver, Colorado.

LENSKY, I. and ROSENFELD, D. (2008). Clouds-aerosols-precipitation satellite analysis tool (CAPSAT). *Atmospheric Chemistry and Physics* **8** 6739–6753.

LINDGREN, F., RUE, H. and LINDSTRÖM, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 423–498. MR2853727

LUCAS, B. D. and KANADE, T. (1981). An iterative image registration technique with an application to stereo vision. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence* **2** 674–679. Morgan Kaufmann, San Francisco, CA.

MAHOWALD, N. M., BAKER, A. R., BERGAMETTI, G., BROOKS, N., DUCE, R. A., JICKELLS, T. D., KUBILAY, N., PROSPERO, J. M. and TEGEN, I. (2005). Atmospheric global dust cycle and iron inputs to the ocean. *Global Biogeochemical Cycles* **19** GB4025.

MARZBAN, C. and SANDGATHE, S. (2010). Optical flow for verification. *Weather and Forecasting* **25** 1479–1494.

RIVAS-PEREA, P., ROSILES, J. G. and CHACON, M. (2010). Traditional and neural probabilistic multispectral image processing for the dust aerosol detection problem. In *IEEE Southwest Symposium on Image Analysis Interpretation* (*SSIAI*), 2010 169–172. IEEE, New York.

RUE, H. and HELD, L. (2005). *Gaussian Markov Random Fields*: *Theory and Applications*. *Monographs on Statistics and Applied Probability* **104**. Chapman & Hall/CRC, Boca Raton, FL. MR2130347

RUE, H., MARTINO, S. and CHOPIN, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 319–392. MR2649602

SCHEFZIK, R., THORARINSDOTTIR, T. L. and GNEITING, T. (2013). Uncertainty quantification in complex simulation models using ensemble copula coupling. *Statist. Sci.* **28** 616–640. MR3161590

SCHEPANSKI, K., TEGEN, I. and MACKE, A. (2012). Comparison of satellite based observations of Saharan dust source areas. *Remote Sensing of Environment* **123** 90–97.

SCHEPANSKI, K., TEGEN, I., LAURENT, B., HEINOLD, B. and MACKE, A. (2007). A new Saharan dust source activation frequency map derived from MSG-SEVIRI IR-channels. *Geophysical Research Letters* **34** L13401.

SCHMETZ, J., PILI, P., TJEMKES, S., JUST, D., KERKMANN, J., ROTA, S. and RATIER, A. (2002). An introduction to Meteosat Second Generation (MSG). *Bulletin of the American Meteorological Society* **83** 977–992.

SCHNÖRR, C. (1991). Determining optical flow for irregular domains by minimizing quadratic functionals of a certain class. *Int. J. Comput. Vis.* **6** 25–38.

SEEMANN, S. W., BORBAS, E. E., KNUTESON, R. O., STEPHENSON, G. R. and HUANG, H.-L. (2008). 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** 108–123.

SIMONCELLI, E. P., ADELSON, E. H. and HEEGER, D. J. (1991). Probability distributions of optical flow. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 310–315. IEEE, New York.

SIMPSON, D., LINDGREN, F. and RUE, H. (2012). In order to make spatial statistics computationally feasible, we need to forget about the covariance function. *Environmetrics* **23** 65–74. MR2873784

XU, K., WIKLE, C. K. and FOX, N. I. (2005). A kernel-based spatio-temporal dynamical model for nowcasting weather radar reflectivities. *J. Amer. Statist. Assoc.* **100** 1133–1144. MR2236929

F. E. BACHL
DEPARTMENT OF MATHEMATICAL SCIENCES
UNIVERSITY OF BATH
CLAVERTON DOWN
BATH BA2 7AY
UNITED KINGDOM

A. LENKOSKI
T. L. THORARINSDOTTIR
NORWEGIAN COMPUTING CENTER
GAUSTADALLEEN 23A
KRISTEN NYGAARDS HUS
NO-0373 OSLO
NORWAY
E-MAIL: alex@nr.no

C. S. GARBE
IMAGE PROCESSING AND MODELING
INTERDISCIPLINARY CENTER FOR SCIENTIFIC COMPUTING (IWR)
UNIVERSITY OF HEIDELBERG
SPEYERER STRASSE 6
D-69115 HEIDELBERG
GERMANY