2014

Predictive inference for big, spatial, non-Gaussian data: MODIS cloud data and its change-of-support

Aritra Sengupta
The Ohio State University

Noel Cressie
University of Wollongong

Brian H. Kahn
California Institute of Technology

Richard Frey
University of Wisconsin-Madison

Predictive inference for big, spatial, non-Gaussian data: MODIS cloud data and its change-of-support

Abstract

Remote sensing of the earth with satellites yields datasets that can be massive in size, nonstationary in space, and non-Gaussian in distribution. To overcome computational challenges, we make use of the reduced-rank Spatial Random Effec ts (SRE) model in our statistical analysis of cloud mask data from NASA's Moderate Resolut ion Imaging Spectroradiometer (MODIS) instrument on board NASA's Terra satellite, launch ed in December 1999. Clouds are the biggest source of uncertainty in future projections of Earth's climate and explain the wide spread of climate sensitivity calculated by climate models due to their inherent differences in parameterized cloud processes. An accurate quantification of the spatial distributions of clouds, as well as a rigorously estimated pixel-scale clear -sky-probability process, is needed to establish reliable estimates of cloud-distributional changes and trends caused by climate change. In this article, we give a hierarchical spatial statistical modeling approach for a very large spatial dataset of size 2 . 75 million pixels, corresponding to a granule of the MODIS cloud mask, and then we use spatial change-of-support relationships to estimate cloud fraction at coarser resolutions. Our model is non-Gaussian; it postulates a hidden process for the prob- ability of a clear sky that makes use of the SRE model, EM estimation, and optimal (empirical Bayes) spatial prediction of the clear-sky-probability process. Measures of uncertainty of the resulting optimal map are also given.

NIASRA

NATIONAL INSTITUTE FOR APPLIED STATISTICS RESEARCH AUSTRALIA

National Institute for Applied Statistics Research Australia

University of Wollongong

Working Paper

17-14

Predictive Inference for Big, Spatial, Non-Gaussian Data: MODIS Cloud Data and its Change-of-Support

Aritra Sengupta, Noel Cressie, Brian H. Kahn and Richard Frey

National Institute for Applied Statistics Research Australia, University of Wollongong, Wollongong NSW 2522. Phone +61 2 4221 5435, Fax +61 2 4221 4845. Email: anica@uow.edu.au

Predictive Inference for Big, Spatial, Non-Gaussian Data: MODIS Cloud Data and its Change-of-Support

Aritra Sengupta [*], Noel Cressie [*†], Brian H. Kahn [‡], and Richard Frey [§]

Abstract

Remote sensing of the earth with satellites yields datasets that can be massive in size, nonstationary in space, and non-Gaussian in distribution. To overcome computational challenges, we make use of the reduced-rank Spatial Random Effects (SRE) model in our statistical analysis of cloud mask data from NASA's Moderate Resolution Imaging Spectroradiometer (MODIS) instrument on board NASA's Terra satellite, launched in December 1999. Clouds are the biggest source of uncertainty in future projections of Earth's climate and explain the wide spread of climate sensitivity calculated by climate models due to their inherent differences in parameterized cloud processes. An accurate quantification of the spatial distributions of clouds, as well as a rigorously estimated pixel-scale clear-sky-probability process, is needed to establish reliable estimates of cloud-distributional changes and trends caused by climate change. In this article, we give a hierarchical spatial statistical modeling approach for a very large spatial dataset of size 2.75 million pixels, corresponding to a granule of the MODIS cloud mask, and then we use spatial change-of-support relationships to estimate cloud fraction at coarser resolutions. Our model is non-Gaussian; it postulates a hidden process for the probability of a clear sky that makes use of the SRE model, EM estimation, and optimal (empirical Bayes) spatial prediction of the clear-sky-probability process. Measures of uncertainty of the resulting optimal map are also given.

[*]Department of Statistics, The Ohio State University
[†]National Institute for Applied Statistics Research Australia, University of Wollongong, Australia
[‡]Jet Propulsion Laboratory, California Institute of Technology
[§]Cooperative Institute for Meteorological Satellite Studies, University of Wisconsin-Madison

Keywords: EM estimation; empirical Bayes; MCMC; optimal spatial prediction; SRE model; spatial GLMM

1 Introduction

Clouds are generally characterized by higher reflectances and lower temperatures than the Earth's surface (Ackerman et al., 2010). They are composed of liquid droplets or ice crystals, and they modulate the planet's radiation budget and directly affect climate. They must be accurately described in order to properly assess climatic processes and climate change. The accuracy of remote sensing retrievals can be affected by cloud-contamination of the atmospheric column. Retrievals of certain atmospheric quantities require clear skies (e.g., land-surface properties or atmospheric aerosols, etc.), while retrieval of cloud properties are relevant for cloudy scenes. So it is important to know if an atmospheric column is cloudy or if it is clear. The Moderate Resolution Imaging Spectroradiometer (MODIS) instrument offers the opportunity for multispectral approaches to cloud detection.

Our interest is in the MODIS instrument that is on board the Terra satellite, which was launched by NASA in December 1999. The Level-2 MODIS cloud mask product (Platnick et al., 2003) is produced for pixel arrays at a spatial resolution of 1 km$\times$1 km. Each MODIS cloud-mask-product file covers data collected over a five-minute time interval, and it is called a granule. Hence, each granule represents a big dataset of approximately 2.75 million pixels of 1 km^2 resolution. In this article, a particular granule of Terra MODIS data (corresponding to June 29, 2006, 12:45 UTC) will be used to illustrate our statistical-modeling approach; an RGB image of the granule is shown in Figure 1.

—— Figure 1 approximately here ——

The MODIS instrument collects data on spectral radiances that are then processed using the MODIS cloud-detection algorithm (e.g., Platnick et al., 2003; Ackerman et al., 1998, 2010) to produce a Level-2 cloud-mask classification (MOD06 product). The MODIS cloud-detection algorithm is based on a number of spectral tests; different tests can have different results for a particular

pixel. The results from all tests are then combined to determine an overall "confidence" value for a pixel, located at $\mathbf{s}$, to be clear (i.e., cloud-free), which we call the "initial Q-values," and we denote it by $Q_I(\mathbf{s})$. Now $0 \leq Q_I(\cdot) \leq 1$: if $Q_I(\mathbf{s})$ is near 1, there is high confidence that the pixel located at $\mathbf{s}$ is clear; and if $Q_I(\mathbf{s})$ is near 0, there is high confidence that the pixel is cloudy. Subsequently, "clear-sky restoral" tests are performed that check for unambiguous clear-sky signals.

The clear-sky restoral tests alter the initial Q-values and result in what we call the "final Q-values," which we denote by $Q(\cdot)$. Again, $0 \leq Q(\cdot) \leq 1$. See Figure 2 for a visual impression of the difference between initial and final Q-values for the granule shown in Figure 1. Noticeable in Figure 1 is a strip of sun-glint reflecting off the ocean, which appears in the left panel of Figure 2 (initial Q-values) but not in the right panel (final Q-values). Thus, restoral tests are important, since there are situations where the cloud-mask algorithm tends to over-predict clouds (e.g., in regions with sun-glint). In what follows, we refer to the final Q-values (which recall are denoted by $Q(\cdot)$), as the *MODIS cloud data*. Again, $Q(\mathbf{s})$ near 1 signifies high confidence for the pixel located at $\mathbf{s}$ to be clear, and $Q(\mathbf{s})$ near 0 signifies high confidence for that pixel to be cloudy. For a basic description of the MODIS cloud-mask algorithm, see Sengupta et al. (2012).

—— Figure 2 approximately here ——

Sengupta et al. (2012) also sketched out a hierarchical spatial statistical model for analyzing MODIS cloud data; here we give full details, including uncertainty quantification, of the nonstationary, non-Gaussian hierarchical spatial statistical model that we fit to the 2.75 million MODIS cloud data shown in the right panel of Figure 2. We also discuss spatial change-of-support, which is important for answering geophysical questions about clouds at resolutions other than that of the data.

The data are defined on 1-km$\times$1-km pixels; henceforth, each of these pixels will be called a "basic areal unit" (BAU). The precise number of BAUs in the granule shown in Figure 1 is $N = 2,748,620$. In general, we assume that data are available for n BAUs, where $n \leq N$. For the particular granule that we consider in this article, $n = N$ (i.e., there are no BAUs without data). A full-rank spatial-statistical modeling approach for the granule would require specifying an

$N \times N$ covariance matrix for the underlying spatial (transformed) clear-sky-probability process. To produce optimal spatial-statistical predictions, we would need to invert $N \times N$ covariance matrices, something that is not computationally feasible for N larger than several thousand.

The computational bottleneck that arises due to the computational cost of inverting the $N \times N$ covariance matrix referred to above, is often referred to as a "big N" problem. When the data appear to be Gaussian, reduced-rank-modeling approaches have been developed to deal with this computational challenge (e.g., Wikle and Cressie, 1999; Wikle et al., 2001; Cressie and Johannesson, 2006, 2008; Banerjee et al., 2008; Stein, 2008; Lopes et al., 2008). For data appearing to come from the exponential family of distributions, Lopes et al. (2011) took the hierarchical generalized linear mixed modeling framework proposed by Diggle et al. (1998) and introduced a new class of models using a latent factor-analysis structure; their fully Bayesian model allows for dimension reduction and hence fast computations. A number of spatial and spatio-temporal applications for very-large-to-massive datasets center around these reduced-rank representations of a hidden continuous Gaussian process (e.g., see the review in Wikle, 2010).

In contrast to the geostatistical-model-based approaches discussed above, a Gaussian Markov random field (GMRF) captures the spatial dependence in the underlying latent process through the (typically sparse) precision matrix, which is the inverse of the covariance matrix. A detailed discussion of this can be found in Rue and Held (2005, Chapter 5) and Cressie and Wikle (2011, Pages 185-186). Rue and Held (2005, Chapter 5) discuss a way to approximate a geostatistical model with a sparse conditional autoregressive spatial model, and this relationship has been used by Lindgren et al. (2011) and Simpson et al. (2012) to build hierarchical spatial models with GMRF process models that allow fast computations. However, the desire for computational efficiency means that they use only a small number of parameters, which could be problematic when modeling spatial dependence over large, continental-scale, heterogeneous regions. Hughes and Haran (2012) consider a Bayesian hierarchical model with a hidden GMRF and use a dimension-reduction approach to deal with spatial confounding and computational complexity that arise when analyzing a large spatial dataset. They parameterize the precision matrix using an underlying graph, where edges of the graph represent spatial dependence, and they assume only a small number of parameters.

To solve the "big N" problem that arises in our application, we shall use the reduced-rank-modeling approach developed by Cressie and Johannesson (2006, 2008), although the MODIS cloud data are bimodal and constrained to $[0, 1]$. Our methodology is a combination of the GLMM framework of Diggle et al. (1998) and the Spatial Random Effects (SRE) model of Cressie and Johannesson (2006, 2008), there developed for Gaussian data with a continuous spatial index. We take an empirical hierarchical modeling (EHM) approach and hence, unlike in a Bayesian hierarchical modeling (BHM) approach where specification of priors is needed, we treat the model's parameters as fixed but unknown. We estimate these parameters using an EM algorithm (e.g., Dempster et al., 1977). Computations for optimal spatial prediction, which includes parameter estimation, are $O(N)$ and feasible. For a more complete discussion of the EHM and BHM approaches, see Cressie and Wikle (2011, Chapter 2).

Cressie and Johannesson (2006, 2008) developed the Spatial Random Effects (SRE) model for optimal spatial predictions from continuous, symmetric data with a continuous spatial index, a methodology that is known as Fixed Rank Kriging (FRK). Cressie and Johannesson (2008) took an EHM approach and gave a method-of-moments estimator of the parameters of the SRE model; Katzfuss and Cressie (2009) took the same EHM approach but developed an EM algorithm to obtain a maximum likelihood (ML) estimates of the parameters; and a Bayesian version is given in Kang and Cressie (2011). The applications in these articles focus on the analysis of very-large-to-massive remote sensing datasets (or some transformation of the data) that appear to come from the Gaussian distribution. In Sengupta and Cressie (2013a,b), a hierarchical spatial statistical model was developed for big, spatial, discrete or continuous data, that includes the SRE model as a component of the process model. Sengupta and Cressie (2013b) used the SRE model in a hierarchical framework to analyze highly skewed, non-negative, remotely sensed Aerosol Optical Depth data, where the models were non-Gaussian and non-additive.

The plan for the rest of the article is as follows: Details of the hierarchical spatial statistical model will be reviewed in Section 2. In Section 3, we outline the statistical inference, which is based on generating MCMC samples from the predictive distribution. The EM algorithm for obtaining the ML estimates of the model parameters is discussed in Section 4. Then, in Section

5, we analyze the granule of MODIS cloud data shown in Figure 2, using the methodologies detailed in this article. In Section 6, we discuss the change-of-support relationship and use it to obtain estimates for cloud fraction at resolutions coarser than the BAU. Discussion and conclusions follow in Section 7.

2 Hierarchical Model for the MODIS Cloud Data

In this section, we give details on the empirical hierarchical model for final Q-values obtained from the MODIS cloud mask product that was proposed in an unrefereed conference proceedings (Sengupta et al., 2012). We index the set of BAUs that have data associated with them as $D_O \equiv \{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$, and the complimentary set of BAUs without data is $D_U \equiv \{\mathbf{s}_{n+1}, \ldots, \mathbf{s}_N\}$. Hence, our data are $\{Q(\mathbf{s}_i) : \mathbf{s}_i \in D_O\}$. For the granule shown in Figure 1, we have $n = N = 2,748,620$ (i.e., no missing data), and $\{Q(\mathbf{s}_i) : \mathbf{s}_i \in D_O\}$ are shown in the right panel of Figure 2.

We introduce a hidden variable $W(\mathbf{s}_i)$ that denotes the true state of the pixel located at $\mathbf{s}_i$ ($i = 1, \ldots, N$), namely 0 (i.e., cloudy) or 1 (i.e., clear). Then we assume a hidden spatial Gaussian process $Y(\cdot)$ that controls the probability of $W(\cdot)$ being 1, where both $W(\cdot)$ and $Y(\cdot)$ are defined over the entire spatial domain, $D \equiv D_O \cup D_U$. That is, $W(\cdot)$ is a binary spatial process whose probability distribution is controlled by a nonlinear transformation of a Gaussian process $Y(\cdot)$; our goal is to make inference on the clear-sky-probability process, $\{P(W(\mathbf{s}) = 1) : \mathbf{s} \in D\}$ from the data $\{Q(\mathbf{s}_i) : \mathbf{s}_i \in D_O\}$.

Our hierarchical spatial statistical model consists of a *data model* and a *two-stage process model*, for which we now give details. In what follows, $[A]$ denotes the probability distribution of A and $[A|B]$ denotes the conditional probability distribution of A given B.

2.1 Data Model

We assume conditional independence and model the pixel-level conditional probabilities,

$$\{[Q(\mathbf{s}_i)|W(\mathbf{s}_i),\ \textit{parameters}] : i = 1, \ldots, n\},$$

using a "0-1 inflated" Beta distribution. Now, conditional on $W(\mathbf{s}_i) = 0$, $Q(\mathbf{s}_i)$ is modeled using a zero-inflated Beta distribution; and conditional on $W(\mathbf{s}_i) = 1$, $Q(\mathbf{s}_i)$ is modeled using a one-inflated Beta distribution. The 0-1 inflation deals with those $\{Q(\mathbf{s}_i)\}$ that are exactly 0 or 1.

Then our *data model* is as follows: For $\mathbf{s}_i \in D_O$, independently,

$$[Q(\mathbf{s}_i)|W(\mathbf{s}_i) = 0, P_0, \alpha_0] = \Big\{P_0 I(Q(\mathbf{s}_i) = 0) + (1 - P_0) f_{1,\alpha_0}(Q(\mathbf{s}_i))\Big\}; \tag{1}$$

and

$$[Q(\mathbf{s}_i)|W(\mathbf{s}_i) = 1, P_1, \alpha_1] = \Big\{P_1 I(Q(\mathbf{s}_i) = 1) + (1 - P_1) f_{1,\alpha_1}(Q(\mathbf{s}_i))\Big\}. \tag{2}$$

In (1) and (2),

$$f_{a,b}(Q(\mathbf{s}_i)) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} Q(\mathbf{s}_i)^{a-1}(1 - Q(\mathbf{s}_i))^{b-1} I(0 < Q(\mathbf{s}_i) < 1), \tag{3}$$

which is the density of a Beta(a,b) random variable, where $a > 0$ and $b > 0$. That is, the parameters in the data model are P_0, α_0, P_1, and α_1; in this article, we consider them to be fixed, unknown, and to be estimated. It is possible that other granules may require use of a different a, b in the $f_{a,b}(\cdot)$ used in (1) and (2).

2.2 Process Model

Next we specify the *two-stage process model*. Recall that the processes $W(\cdot)$ and $Y(\cdot)$ are defined over the entire spatial domain D, not just over D_O. In what follows, we shall use Gau$(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ to denote a Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

"Process model 1" represents the distribution of $\{W(\mathbf{s}_i) : i = 1, \ldots, N\}$, conditional on the hidden spatial process $Y(\cdot)$. We assume a set of independent Bernoulli random variables for *process model 1*: That is, for $i = 1, \ldots, N$, independently,

$$W(\mathbf{s}_i)|Y(\cdot) \sim \text{Bernoulli}(p(\mathbf{s}_i)), \tag{4}$$

where recall that $W(\mathbf{s}_i) = 1$ (respectively, 0) means that the pixel located at $\mathbf{s}_i$ is clear (respectively, cloudy). Then $p(\cdot) = Pr(W(\cdot) = 1)$ is the clear-sky-probability process, and

$$Y(\cdot) \equiv \log\left(\frac{p(\cdot)}{1-p(\cdot)}\right),$$

is the logit transform of $p(\cdot)$; and conversely,

$$p(\cdot) = \left(\frac{\exp(Y(\cdot))}{1+\exp(Y(\cdot))}\right). \tag{5}$$

At the second stage of the process model ("process model 2"), we use the reduced-rank Spatial Random Effects (SRE) model (e.g., Cressie and Johannesson, 2006, 2008) to define the spatial dependence in $Y(\cdot)$. *Process model 2* is:

$$Y(\mathbf{s}_i) = \mathbf{X}(\mathbf{s}_i)^\top \boldsymbol{\beta} + \mathbf{S}(\mathbf{s}_i)^\top \boldsymbol{\eta} + \xi(\mathbf{s}_i);\ i = 1, \ldots N, \tag{6}$$

where $\mathbf{X}(\mathbf{s}_i)$ is a p-dimensional vector of known covariates at each location $\mathbf{s}_i$; $\boldsymbol{\beta}$ denotes the set of p unknown regression coefficients; $\mathbf{S}(\cdot) \equiv (S_1(\cdot), \ldots, S_r(\cdot))^\top$ is a vector of r (not necessarily orthogonal) spatial basis functions, where r is fixed and much smaller than n; $\boldsymbol{\eta}$ is an r-dimensional vector of spatial random effects assumed to have a Gau$(\mathbf{0}, \mathbf{K})$ distribution; and $\xi(\cdot)$ is a fine-scale-variation process modeled as independent Gau$(0, \sigma^2_\xi)$ random variables. In process model 2, the parameters $\boldsymbol{\beta}$, $\mathbf{K}$, and σ^2_ξ are fixed, unknown, and to be estimated.

3 Empirical Bayesian Inference

Our main focus in this paper is on prediction of the clear-sky-probability process, $p(\cdot)$, given by (5). Consequently, the probability of cloudiness, $q(\cdot)$, can be obtained as:

$$q(\cdot) = 1 - p(\cdot). \tag{7}$$

From data $\{Q(\mathbf{s}_i) : i = 1,\ldots,n\}$ at $n \leq N$ spatial locations $\{\mathbf{s}_1,\ldots,\mathbf{s}_n\}$, we wish to make inference on the hidden process $Y(\cdot)$ (or equivalently $p(\cdot)$) at all spatial locations $\{\mathbf{s}_i : i = 1,\ldots,N\}$. The parameters $\boldsymbol{\theta} \equiv \left\{P_0, P_1, \alpha_0, \alpha_1, \boldsymbol{\beta}, \mathbf{K}, \sigma^2_\xi\right\}$ are unknown and considered to be nuisance parameters for the purpose of prediction. Our hierarchical model becomes an *empirical* hierarchical model when the parameters are estimated from the data instead of a prior distribution being put on them (Section 4). They are then substituted into the predictive distribution of $Y(\cdot)$.

Define $\mathbf{Q}_O \equiv (Q(\mathbf{s}_1),\ldots,Q(\mathbf{s}_n))^\top$, $\mathbf{W} \equiv (W(\mathbf{s}_1),\ldots W(\mathbf{s}_N))^\top$, $\mathbf{Y} \equiv (Y(\mathbf{s}_1),\ldots,Y(\mathbf{s}_N))^\top$, and $\boldsymbol{\xi} \equiv (\xi(\mathbf{s}_1),\ldots \xi(\mathbf{s}_N))^\top$. Then our immediate goal is to predict the process variables, $(\mathbf{W},\mathbf{Y})$, given the data and the parameters. The predictive distribution is:

$$[\mathbf{W},\mathbf{Y}|\mathbf{Q}_O,\boldsymbol{\theta}] \propto [\mathbf{Q}_O|\mathbf{W},\mathbf{Y},\boldsymbol{\theta}][\mathbf{W}|\mathbf{Y},\boldsymbol{\theta}][\mathbf{Y}|\boldsymbol{\theta}]. \tag{8}$$

However, the predictive distribution is not available in closed form, nor is $\boldsymbol{\theta}$ known. To solve these problems, we shall use EM estimation of $\boldsymbol{\theta}$ to yield $\hat{\boldsymbol{\theta}}_{EM}$, and we shall use a Markov Chain Monte Carlo (MCMC) algorithm (see, e.g., Robert and Casella, 2004) to yield samples from the predictive distribution, $[\mathbf{W},\mathbf{Y}|\mathbf{Q}_O,\boldsymbol{\theta}]$, where $\hat{\boldsymbol{\theta}}_{EM}$ is substituted in for $\boldsymbol{\theta}$. From (6), we can see that this is achieved by obtaining samples from the predictive distribution, $[\mathbf{W},\boldsymbol{\eta},\boldsymbol{\xi}|\mathbf{Q}_O,\boldsymbol{\theta}]$. The MCMC algorithm to obtain this predictive distribution for the MODIS cloud data analyzed in Section 5, is described in the Appendix.

4 EM Estimation of Parameters

The EM algorithm (Dempster et al., 1977) has been employed for estimation of parameters in the presence of missing data; for details on the generic EM algorithm, see McLachlan and Krishnan (2008). The EM algorithm involves iterating between an E (expectation) step and an M (maximization) step, and in our case the E-step is the most problematic. When the integrals in the E-step are not available in closed form, one common approach is to implement a stochastic EM (SEM) algorithm (see McLachlan and Krishnan, 2008; Robert and Casella, 2004), where the expectations are

evaluated using Monte Carlo integration. When datasets are massive, as in our case, this computation will be extremely slow, and hence the EM algorithm will be computationally very expensive. Following the ideas in Sengupta and Cressie (2013b), we instead derive Laplace approximations (LA) to approximate the expectations involved in the E-step of the EM algorithm.

At locations in D_O, the process $W(\cdot)$, the random effects $\boldsymbol{\eta}$ and the fine-scale-variation term $\xi(\cdot)$ are not observed and can be treated as missing data. The "complete data" log likelihood, L_c, for the unknown parameters is made up of the observations $\mathbf{Q}_O$ and the "missing data" $\mathbf{W}_O$, $\boldsymbol{\eta}$, and $\boldsymbol{\xi}_O$, where $\mathbf{W}_O \equiv (W(\mathbf{s}_1),\ldots,W(\mathbf{s}_n))^\top$, and $\boldsymbol{\xi}_O \equiv (\xi(\mathbf{s}_1),\ldots,\xi(\mathbf{s}_n))^\top$. Then L_c is simply the logarithm of the joint distribution of $\mathbf{Q}_O$, $\mathbf{W}_O$, $\boldsymbol{\eta}$, and $\boldsymbol{\xi}_O$, given the parameters $\boldsymbol{\theta} = \left\{P_0, \alpha_0, P_1, \alpha_1, \boldsymbol{\beta}, \mathbf{K}, \sigma^2_\xi\right\}$. That is,

$$
\begin{aligned}
&L_c(\boldsymbol{\theta}|\mathbf{Q}_O,\boldsymbol{\eta},\boldsymbol{\xi}_O)\\
&= \log[\mathbf{Q}_O|\mathbf{W}_O,\boldsymbol{\beta},\boldsymbol{\eta},\boldsymbol{\xi}_O] + \log[\mathbf{W}_O|\boldsymbol{\beta},\boldsymbol{\eta},\boldsymbol{\xi}_O] + \log[\boldsymbol{\eta}|\mathbf{K}] + \log[\boldsymbol{\xi}_O|\sigma^2_\xi]\\
&= \text{const.} + \sum_{i:Q(\mathbf{s}_i)=1} W(\mathbf{s}_i)\log P_1 + \sum_{i:Q(\mathbf{s}_i)=0}(1-W(\mathbf{s}_i))\log P_0\\
&\quad + \sum_{i:0<Q(\mathbf{s}_i)<1} W(\mathbf{s}_i)\log\left\{(1-P_1)\,f_{1,\alpha_1}(Q(\mathbf{s}_i))\right\}\\
&\quad + \sum_{i:0<Q(\mathbf{s}_i)<1}(1-W(\mathbf{s}_i))\log\left\{(1-P_0)\,f_{1,\alpha_0}(Q(\mathbf{s}_i))\right\}\\
&\quad + \sum_{i=1}^{n} W(\mathbf{s}_i)(\mathbf{X}(\mathbf{s}_i)^\top\boldsymbol{\beta} + \mathbf{S}(\mathbf{s}_i)^\top\boldsymbol{\eta} + \xi(\mathbf{s}_i))\\
&\quad - \sum_{i=1}^{n}\log\left(1+\exp(\mathbf{X}(\mathbf{s}_i)^\top\boldsymbol{\beta} + \mathbf{S}(\mathbf{s}_i)^\top\boldsymbol{\eta} + \xi(\mathbf{s}_i))\right)\\
&\quad - \frac{1}{2}\log|\mathbf{K}| - \frac{1}{2}\text{trace}\left(\boldsymbol{\eta}\boldsymbol{\eta}^\top\mathbf{K}^{-1}\right)\\
&\quad - \frac{n}{2}\log\sigma^2_\xi - \frac{1}{2\sigma^2_\xi}\text{trace}\left(\boldsymbol{\xi}_O\boldsymbol{\xi}_O^\top\right),
\end{aligned}
\tag{9}
$$

where $f_{a,b}(Q(\cdot))$ is given by (3), and "const." denotes a generic constant that does not depend on the parameters.

Assume that we have completed the l-th iteration of the EM algorithm. We now discuss the E-step and the M-step for the $(l+1)$-th iteration of the EM algorithm. At the $(l+1)$-th

iteration of the EM algorithm, the E-step involves taking the expectation of the complete data loglikelihood with respect to (w.r.t.) the missing data $(\mathbf{W}_O, \boldsymbol{\eta}, \boldsymbol{\xi}_O)$, conditional on the observed data $\mathbf{Q}_O$ and the parameter values, $\boldsymbol{\theta}^{[l]}$, at the l-th iteration. We obtain the expectation iteratively, first w.r.t. $[\mathbf{W}_O|\boldsymbol{\eta}, \boldsymbol{\xi}_O, \mathbf{Q}_O, \boldsymbol{\theta}^{[l]}]$ and then w.r.t. $[\boldsymbol{\eta}, \boldsymbol{\xi}_O|\mathbf{Q}_O, \boldsymbol{\theta}^{[l]}]$. The conditional distribution $[W(\mathbf{s}_i)|\boldsymbol{\eta}, \boldsymbol{\xi}_O, \mathbf{Q}_O, \boldsymbol{\theta}^{[l]}]$ is available in closed form. However, the second expectation, w.r.t. the distribution $[\boldsymbol{\eta}, \boldsymbol{\xi}_O|\mathbf{Q}_O, \boldsymbol{\theta}^{[l]}]$, is not, and we approximate it with a Laplace approximation (e.g., Sengupta and Cressie, 2013a,b) based on a second-order Taylor-series expansion of the logarithm of the integrands around their respective modes.

Following the E-step, we perform the M-step, where we maximize the expectation evaluated at the E-step with respect to each of the parameters in $\boldsymbol{\theta}$: The maximization w.r.t. P_0, P_1, $\mathbf{K}$, and σ^2_ξ is obtained by differentiating the expectation of the complete data loglikelihood given by (9), w.r.t. P_0, P_1, $\mathbf{K}$, and σ^2_ξ, and then equating the result to zero. Solving these equations yields:

$$
\begin{aligned}
\sigma^{2[l+1]}_\xi &= \frac{1}{n}\text{trace}\left(E(\boldsymbol{\xi}_O|\mathbf{Q}_O, \boldsymbol{\theta}^{[l]})E(\boldsymbol{\xi}_O|\mathbf{Q}_O, \boldsymbol{\theta}^{[l]})^\top + \text{var}\left(\boldsymbol{\xi}_O|\mathbf{Q}_O, \boldsymbol{\theta}^{[l]}\right)\right) \\
\mathbf{K}^{[l+1]} &= E(\boldsymbol{\eta}|\mathbf{Q}_O, \boldsymbol{\theta}^{[l]})E(\boldsymbol{\eta}|\mathbf{Q}_O, \boldsymbol{\theta}^{[l]})^\top + \text{var}\left(\boldsymbol{\eta}|\mathbf{Q}_O, \boldsymbol{\theta}^{[l]}\right) \\
P_0^{[l+1]} &= \frac{\sum_{i:Q(\mathbf{s}_i)=0}\left(1 - E(W(\mathbf{s}_i)|\mathbf{Q}_O, \boldsymbol{\theta}^{[l]})\right)}{\sum_{i:Q(\mathbf{s}_i)<1}\left(1 - E(W(\mathbf{s}_i)|\mathbf{Q}_O, \boldsymbol{\theta}^{[l]})\right)} \\
P_1^{[l+1]} &= \frac{\sum_{i:Q(\mathbf{s}_i)=1} E(W(\mathbf{s}_i)|\mathbf{Q}_O, \boldsymbol{\theta}^{[l]})}{\sum_{i:Q(\mathbf{s}_i)>0} E(W(\mathbf{s}_i)|\mathbf{Q}_O, \boldsymbol{\theta}^{[l]})}.
\end{aligned} \tag{10}
$$

The maximizations with respect to α_0, α_1, and $\boldsymbol{\beta}$ are not available in closed form. We use a one-step Newton-Raphson update for these parameters, within each iteration of the EM algorithm. For complete technical details of the EM algorithm, including the Laplace approximation and the choice of starting values, see Sengupta (2012, Section 4.5).

5 Spatial Statistical Analysis of MODIS Cloud Data

In this section, we carry out a spatial statistical analysis of the $n = 2,748,620$ MODIS cloud data shown in the right panel of Figure 2, using the hierarchical spatial statistical model discussed in

Section 2. We use it to predict clear-sky probabilities for the entire study region, $D = D_O \cup D_U$, along with estimates of prediction uncertainties. A map showing the geographical location of the granule is given in Figure 3; the study region lies in a part of the globe from longitudes -45 degrees to -10 degrees, and from latitudes 20 degrees to 45 degrees. Recall that the granule was collected by the MODIS instrument on the Terra satellite and corresponds to June 29, 2006, 12:45 UTC. The resolution of the data is 1 km$\times$1 km and, since we have data for all the locations, $n = N = 2,748,620$.

—— Figure 3 approximately here ——

5.1 Spatial Basis Functions

In this section, we discuss the choice of spatial basis function. The selection of the basis functions, its type and number, is a current area of research (e.g., Bradley et al., 2011). For the purpose of this analysis, we selected as basis functions the bisquare functions (e.g., Cressie and Johannesson, 2006, 2008). The generic form of a bisquare function is,

$$b(\mathbf{s}) = \left\{ 1 - \left(\frac{||\mathbf{s} - \mathbf{c}||}{w} \right)^2 \right\}^2 I(||\mathbf{s} - \mathbf{c}|| < w), \tag{11}$$

where $\mathbf{c}$ is the center of the basis function, and $I(A)$ is an indicator function equal to 1 if A is true, and equal to 0 otherwise. Basis-function centers $\{\mathbf{c}_j\}$ in D are usually chosen according to a multi-resolution scheme (e.g., a quad-tree). Finally, the "aperture" w given by Cressie and Johannesson (2008) is,

$$w = 1.5 \times \text{shortest great-arc distance between like-resolution center points.}$$

An illustration of a generic bisquare basis function is given in Figure 4. Other choices for basis functions are also possible (e.g., EOFs in Wikle and Cressie, 1999; eigenfunctions in Wikle et al., 2001; W-wavelets in Shi and Cressie, 2007).

—— Figure 4 approximately here ——

We employ several resolutions of basis functions to capture the different scales of spatial variability; here we use three resolutions to obtain $\{b_j(\mathbf{s}) : j = 1, \ldots, (r_1 + r_2 + r_3)\}$, where $r_1 = 12$, $r_2 = 34$, and $r_3 = 102$, are the number of basis functions at the three resolutions. The centers of the bisquare basis functions were selected using a quad-tree structure (e.g., Cressie and Kang, 2010), ensuring that the centers for the different resolutions do not match. The number of basis functions were determined to allow full coverage of the spatial domain. We also included centers of the bisquare function outside the study region to account for boundary effects (e.g., Cressie and Kang, 2010). We further standardized the bisquare function $b_j(\cdot)$ to obtain the j-th basis function,

$$S_j(\mathbf{s}) \equiv \frac{b_j(\mathbf{s}) - \text{ave}_{\mathbf{s}\in D}\left(b_j(\mathbf{s})\right)}{\left\{\text{var}_{\mathbf{s}\in D}\left(b_j(\mathbf{s})\right)\right\}^{1/2}}; \; j = 1, \ldots, (r_1 + r_2 + r_3), \tag{12}$$

where $\text{ave}_{\mathbf{s}\in D}(\cdot)$ and $\text{var}_{\mathbf{s}\in D}(\cdot)$ are spatial moments taken over the domain of interest, and $\mathbf{s} \in D$. The locations of the basis-function centers for all three resolutions are shown in Figure 5.

—— Figure 5 approximately here ——

5.2 Optimal Spatial Mapping of the Clear-Sky-Probability Process

We now discuss the results obtained by fitting the hierarchical statistical model presented in Section 2, to the MODIS cloud data. Subsequently, we produce optimal spatial maps showing the predictions for the underlying clear-sky-probability process, along with maps showing the prediction uncertainties.

First, let us consider the covariates $\mathbf{X}(\cdot)$ in (6). We include the vector $\mathbf{1}$ and a vector of latitudes as covariates. Further, instead of using the $r_1 = 12$ coarsest-resolution bisquare functions as spatial basis functions, we use them as a further 12 covariates in $\mathbf{X}(\cdot)$ (e.g., Shi and Cressie, 2007).

The second term of (6) involves an r-dimensional vector, $\mathbf{S}(\cdot)$, of spatial basis functions, which includes the $r_2 + r_3 = 136$ bisquare functions at the second and the third resolutions (see Figure 5). Now, there are regions in the study region that are affected by sun-glint (see Figure 1), which the MODIS cloud algorithm attempts to account for by implementing clear-sky-restoral tests. Nevertheless, the presence or absence of sun-glint remains a source of spatial variability for the granule

we consider. Hence, we include the sun-glint indicator flag (which takes a value 1 if a pixel is affected by sun-glint, and is 0 otherwise, and is available as part of the MODIS cloud product) as a basis function in the vector of basis functions $\mathbf{S}(\cdot)$. That is, $r = 1 + r_2 + r_3 = 1 + 34 + 102 = 137$.

With the model specified as above, we implemented the EM algorithm to obtain estimates of the parameters $\boldsymbol{\theta}$ defined in Section 2. The EM algorithm converged after 14 iterations, and the computational time for the EM algorithm was 27.76 minutes.

After having obtained the EM estimates, $\hat{\boldsymbol{\theta}}_{EM} \equiv \left\{\hat{P}_{0;EM}, \hat{P}_{1;EM}, \hat{\alpha}_{0;EM}, \hat{\alpha}_{1;EM}, \hat{\boldsymbol{\beta}}_{EM}, \hat{\mathbf{K}}_{EM}, \hat{\sigma}^2_{\xi;EM}\right\}$, we substituted them into the MCMC algorithm (see Appendix) to obtain samples from the (empirical) predictive distribution, $[\mathbf{W}, \boldsymbol{\eta}, \boldsymbol{\xi} | \mathbf{Q}_O, \hat{\boldsymbol{\theta}}_{EM}]$. We generated 10,000 MCMC samples, after discarding 1,000 samples as burn-in. The computational time for the MCMC was 12.73 hours. (All computations were performed on a dual quad core 2.8 GHz 2x Xeon X5560 processor, with 96 Gbytes of memory.)

The hierarchical nature of the model allows us to look at the separate sources of variability, on the logit scale (i.e., we can separate out the components of $Y(\cdot)$). The predictive mean of these different sources of variability, along with that of their sum $Y(\cdot)$, are shown in Figure 6.

—— Figure 6 approximately here ——

Now, using the MCMC samples referred to above, we computed the predictive mean and the predictive standard deviation of the clear-sky probability $p(\mathbf{s}_i)$ given by (5), for $i = 1, \ldots, N$. We also obtained the pixelwise 2.5 and 97.5 percentiles of each of the N elements of $\mathbf{p} \equiv (p(\mathbf{s}_1), \ldots, p(\mathbf{s}_N))^\top$ from the predictive distribution, $[\mathbf{p} | \mathbf{Q}_O, \hat{\boldsymbol{\theta}}_{EM}]$, computed from $[\mathbf{W}, \boldsymbol{\eta}, \boldsymbol{\xi} | \mathbf{Q}_O, \hat{\boldsymbol{\theta}}_{EM}]$. Figure 7 shows maps of the pixelwise predictive mean, the pixelwise predictive standard deviation, and the pixelwise 2.5 and 97.5 percentiles, respectively; the latter two quantities are the lower and upper end-points of a pixelwise 95% prediction interval.

—— Figure 7 approximately here ——

6 Change-of-Support

In this section, we discuss spatial change-of-support and outline the role it plays in predicting cloud fraction at resolutions coarser than the base resolution of 1 km$\times$1 km. Our goal is to demonstrate that, by incorporating spatial statistical models into the problem, optimal estimates with much higher accuracy and precision, can be obtained.

We start by defining cloud fraction and the current, straightforward way it is estimated. Then we propose an alternative estimator, specifically an optimal Bayes (OB) estimator based on the spatial hierarchical model defined in Section 2, and we show that it minimizes the mean squared error incurred when estimating the cloud fraction. This assumes that all the parameters are known; by substituting parameter estimates into the OB estimator, we obtain an *empirical Bayes (EB) estimator* that we show, via a simulation study, is much more efficient than the straightforward estimator. Hence, we recommend this new, EB estimator of cloud fraction. Finally, we look at a geophysical question involving cloud fraction at 5 km$\times$5 km resolutions and compare the results obtained using the straightforward estimator and the EB estimator.

6.1 Estimation of Cloud Fraction Using a Straightforward Estimator

In applications such as weather forecasting and climate projections, researchers are interested in geophysical questions that involve cloud fraction (e.g., Di Girolamo and Davies, 1997). Cloud fraction is defined as follows: Let $C(\mathbf{s}) = 1$ if there is cloud present (and 0 otherwise), where $\mathbf{s} \in D$, and for D a continuous-spatial-index set. Consider an area $B \subset D$. Then the cloud fraction for B is defined as $\int_B C(\mathbf{s})d\mathbf{s}/|B|$, where $|B|$ denotes the area of B. Henceforth, we call B the *spatial support* of the cloud fraction. Now recall that D (and hence B) has been discretized into 1 km$\times$1 km BAUs; then the cloud fraction for B is numerically approximated as,

$$CF(B) \equiv \frac{\sum_{\mathbf{s}_i \in B} I\{W(\mathbf{s}_i) = 0\}}{\sum_{\mathbf{s}_i \in B} 1}, \tag{13}$$

where recall from Section 2.2 that $W(\mathbf{s}_i) = 0$ corresponds to the i-th 1 km$\times$1 km pixel located at $\mathbf{s}_i$ being cloudy; see Zhao and Di Girolamo (2006).

Currently, $CF(B)$ is estimated in the following straightforward way: Classify each 1 km$\times$1 km pixel as cloudy or clear by thresholding the final Q-values. If $Q(\mathbf{s}_i) > 0.95$, then the pixel $\mathbf{s}_i$ is classified as clear (i.e., $W(\mathbf{s}_i) = 1$), and if $Q(\mathbf{s}_i) \leq 0.95$, then the pixel $\mathbf{s}_i$ is classified as cloudy (i.e., $W(\mathbf{s}_i) = 0$). Then the straightforward (method-of-moments, plug-in) estimate of the cloud fraction over B, henceforth referred to as the *plug-in (PI) estimate*, is given by,

$$CF^{PI}(B) \equiv \frac{\sum_{\mathbf{s}_i \in B} I\{Q(\mathbf{s}_i) \leq 0.95\}}{\sum_{\mathbf{s}_i \in B} 1}. \tag{14}$$

Notice that the estimate given by (14) depends on the threshold value 0.95 used for cloud-mask classification, making it a biased estimator due to classification error. More formally,

$$E(CF^{PI}(B)|\boldsymbol{\theta}) = \sum_{\mathbf{s}_i \in B} P(Q(\mathbf{s}_i) \leq 0.95|\boldsymbol{\theta})/n(B) \neq E(CF(B)|\boldsymbol{\theta}), \tag{15}$$

where $n(B) \equiv \sum_{\mathbf{s}_i \in B} 1$, and $\boldsymbol{\theta}$ here is a vector of fixed but unknown parameters that govern the probability distribution of $W(\cdot)$ in (13). In what follows, we shall show that the PI estimator, as well as being biased, also lacks efficiency when compared to the OB or EB estimators (optimal or near-optimal in terms of minimizing mean squared error).

6.2 Optimal Bayes (OB) Estimator for Cloud Fraction

With the hierarchical model proposed in Section 2, and assuming $\boldsymbol{\theta}$ is known, the OB estimator minimizes the mean squared error and is given by,

$$CF^{OB}(B) \equiv E(CF(B)|\mathbf{Q}_O, \boldsymbol{\theta}) = \sum_{\mathbf{s}_i \in B} P(W(\mathbf{s}_i) = 0|\mathbf{Q}_O, \boldsymbol{\theta})/n(B), \tag{16}$$

where recall that $\mathbf{Q}_O$ are the final Q-values used to classify their respective pixels. Taking the expectation of $CF^{OB}(B)$ with respect to the distribution $[\mathbf{Q}_O|\boldsymbol{\theta}]$, we obtain

$$E(CF^{OB}(B)|\boldsymbol{\theta}) = E\{E(CF(B)|\mathbf{Q}_O,\boldsymbol{\theta})\} = E(CF(B)|\boldsymbol{\theta}). \tag{17}$$

Consequently, the estimation error of the OB estimator, $CF^{OB}(B) - CF(B)$, has mean zero; that is, $CF^{OB}(B)$ is unbiased.

Next we prove that $CF^{OB}(\cdot)$ minimizes the mean squared error among the class of all estimators for $CF(\cdot)$. Consider an arbitrary estimator, $\tilde{CF}(\cdot)$, whose squared error of estimation is:

$$L(CF(B),\tilde{CF}(B)) = (CF(B) - \tilde{CF}(B))^2, \tag{18}$$

where $L(\cdot,\cdot)$ is generic notation for a loss function (e.g., Berger, 1985). Equation (18) emphasizes that we are considering squared-error loss. Conditional on $\boldsymbol{\theta}$, the mean squared error (MSE) is,

$$\begin{aligned} E\left\{(CF(B) - \tilde{CF}(B))^2|\boldsymbol{\theta}\right\} =& E\left\{(CF(B) - CF^{OB}(B) + CF^{OB}(B) - \tilde{CF}(B))^2|\boldsymbol{\theta}\right\} \\ =& E\left\{(CF(B) - CF^{OB}(B))^2|\boldsymbol{\theta}\right\} \\ &+ E\left\{(CF^{OB}(B) - \tilde{CF}(B))^2|\boldsymbol{\theta}\right\} \\ &+ 2E\left\{(CF(B) - CF^{OB}(B))(CF^{OB}(B) - \tilde{CF}(B))|\boldsymbol{\theta}\right\}. \end{aligned} \tag{19}$$

Note that the last term in (19) is equal to zero because $CF^{OB}(B) \equiv E(CF(B)|\mathbf{Q}_O,\boldsymbol{\theta})$. Consequently, the MSEs satisfy,

$$E\left\{\left(CF(B) - \tilde{CF}(B)\right)^2|\boldsymbol{\theta}\right\} \geq E\left\{\left(CF(B) - CF^{OB}(B)\right)^2|\boldsymbol{\theta}\right\}; \tag{20}$$

that is, for squared error loss given by (18), $CF^{OB}(B)$ is the optimal Bayes predictor of $CF(B)$, for any spatial support $B \subset D$.

In practice, $\boldsymbol{\theta}$ is unknown and hence $CF(B)$ is not a function only of the data. The OB estimate

becomes an empirical Bayes (EB) estimate by replacing $\boldsymbol{\theta}$ in (16) with $\hat{\boldsymbol{\theta}}_{EM}$ (see Section 4):

$$CF^{EB}(\cdot) \equiv E(CF(B)|\mathbf{Q}_O, \hat{\boldsymbol{\theta}}_{EM}) = \sum_{\mathbf{s}_i \in B} P(W(\mathbf{s}_i) = 0|\mathbf{Q}_O, \hat{\boldsymbol{\theta}}_{EM})/n(B). \tag{21}$$

The use of $\hat{\boldsymbol{\theta}}_{EM}$ in place of $\boldsymbol{\theta}$ in (21) could introduce bias and typically increases the mean squared error. In the next subsection, we use simulation to demonstrate that these effects are very small and that $CF^{EB}(\cdot)$ given by (21) is much more efficient than $CF^{PI}(\cdot)$ given by (14).

6.3 Simulation Study Comparing Estimators' Bias and Mean Squared Error

In this section, we present the results of a simulation study that first compares the predictions obtained using the OB estimator given by (16) to those obtained using the EB estimator given by (21), and we demonstrate that using $\hat{\boldsymbol{\theta}}_{EM}$ in place of $\boldsymbol{\theta}$ scarcely impacts the bias and mean squared error. Then we compare the PI estimator given by (14) to the EB estimator given by (21).

In this simulation, we consider a smaller-dimensional sub-granule of the usual granule, where $N = n = 625,000$ 1 km$\times$1 km pixels. For context, the original data in the sub-granule, $\mathbf{Q}_O$, are shown in the left panel of Figure 8. In our simulation study, we generate L realizations of $\mathbf{Q}_O$, each of which is subjected to a computationally intensive estimation procedure. Hence, reducing N allows L to be large.

Our simulation study considers several factors. The "resolution factor" has three levels: From the spatial model given in Section 2, the levels are: (a) 4 resolutions (b) the 3 coarsest resolutions selected from the 4 resolutions, and (c) the 3 finest resolutions selected from the 4 resolutions. The centers of the basis function considered at the different resolutions, along with the study domain, are shown in the right panel of Figure 8. (Some centers resulted in a column of zeros in the spatial basis matrix $\mathbf{S}$; for computational efficiency, these were removed from consideration.) The "support factor" is considered at 20 levels, namely B at $2\times 2,\ 4\times 4, \ldots, 40\times 40$ km support. The "treatment factor" has three levels given by the type of estimator used to estimate $CF(\cdot)$. They are: (a) PI estimator given by (14), (b) OB estimator given by (16), and (c) EB estimator given by (21).

Our goal is to compare the estimators and to see how resolution-choice and support affects estimation. The three simulation models corresponding to the resolution choices are calibrated to the data $\mathbf{Q}_O$ shown in the left panel of Figure 8. Our approach here is analogous to a parametric bootstrap (e.g., Efron and Tibshirani, 1993, Section 6.5), where we simulate realizations from a model estimated from the observed data. Our "response" in the simulation study is *mean squared error (MSE)* and *bias*, computed for the three estimators at each of the three levels of the resolution factor and the 20 levels of the support factor.

—— Figure 8 here. ——

Fix a resolution level. The bias for the spatial support B, when using estimator A, is defined as:

$$bias^A(B) \equiv E\left(CF^A(B) - CF(B)|\boldsymbol{\theta}\right). \tag{22}$$

The MSE for the spatial support B, when using estimator A, is defined as:

$$MSE^A(B) \equiv E\left(CF^A(B) - CF(B)|\boldsymbol{\theta}\right)^2. \tag{23}$$

If the estimator A is unbiased, then $MSE^A(B)$ is the same as $V^A(B) \equiv \text{var}\left(CF^A(B) - CF(B)|\boldsymbol{\theta}\right)$.

The quantities (22) and (23) are obtained empirically from the simulation, avoiding the need for any analytic derivations. Analogous to the simulation experiment in Cressie (2006), we empirically obtain $MSE^A(B)$, and $bias^A(B)$ for an estimator $A \in \{PI,\ OB,\ EB\}$ and for support $B \in \{2\times 2, 4\times 4, \ldots 40\times 40\}$, centered on the center of the study domain shown in Figure 8, and nested. These computations are carried out for each of the three levels of the resolution factor.

To compute the first two moments empirically for each level of the resolution factor, we simulate data $\{\mathbf{Q}_O^{[l]}; l = 1,\ldots,L\}$ using the hierarchical-model sequence:

$$[\mathbf{Y}^{[l]}|\boldsymbol{\theta}] \rightarrow [\mathbf{W}^{[l]}|\mathbf{Y}^{[l]}] \rightarrow [\mathbf{Q}_O^{[l]}|\mathbf{W}^{[l]}].$$

Then, from $\mathbf{Q}_O^{[l]}$, we compute $CF^{PI[l]}(B)$, $CF^{OB[l]}(B)$, and $CF^{EB[l]}(B)$. Note that computing the

latter two estimates, $CF^{OB[l]}(B)$ and $CF^{EB[l]}(B)$, requires MCMC-based simulations from the distributions, $[\mathbf{W},\mathbf{Y}|\mathbf{Q}_O^{[l]},\boldsymbol{\theta}]$ and $[\mathbf{W},\mathbf{Y}|\mathbf{Q}_O^{[l]},\hat{\boldsymbol{\theta}}_{EM}]$, respectively. We set $L = 500$, which guarantees accuracy up to the first decimal place (e.g., Aldworth and Cressie, 1999).

6.3.1 Results from the simulation study

We compute the "bias" responses, $bias^{PI}(B)$, $bias^{OB}(B)$, and $bias^{EB}(B)$, for $B \in \{2\times 2, 4\times 4, \ldots 40\times 40\}$, centered on the center of the study domain. They are obtained empirically as follows. When using estimator A, the empirical bias is:

$$\hat{bias}^{A}(B) \equiv (1/L)\sum_{l=1}^{L}\left(CF^{A[l]}(B) - CF^{truth[l]}(B)\right), \tag{24}$$

where $CF^{truth[l]}(B)$ is given by:

$$CF^{truth[l]}(B) \equiv \frac{\sum_{\mathbf{s}_i\in B} I\left\{W^{[l]}(\mathbf{s}_i) = 0\right\}}{\sum_{\mathbf{s}_i\in B} 1}. \tag{25}$$

Figure 9 shows the bias for the PI estimator (left panel) and the EB estimator (right panel), as a function of B. A plot showing the bias for the OB estimator looks almost identical to the one for the EB estimator, and hence it is not shown. Based on Figure 9, we conclude that the EB estimator is approximately unbiased, but that the PI estimator has considerable positive bias.

The PI estimator is based on thresholding $Q(\cdot)$ independently for each 1 km$\times$1 km pixel, where recall that $Q(\cdot) \leq 0.95$ is classified as cloudy and $Q(\cdot) > 0.95$ is classified as clear. The thresholding value of 0.95 biases the classification towards cloudy scenes, which results in overpredictions for $CF(\cdot)$ and leads to the positive bias seen in Figure 9.

From Figure 9, we see that the bias of the PI estimator is larger for the model with four resolutions, when compared to the two models with three resolutions. Now, thresholding $Q(\cdot)$ independently for all the 1 km$\times$1 km pixels ignores the spatial structure. Further, when we have four resolutions in the spatial model, the spatial dependence is stronger and the consequences of ignoring it are greater. The result is more bias in the spatial model with four resolutions.

—— Figure 9 here. ——

Next, we look at the "MSE" responses, $MSE^{PI}(B)$, $MSE^{OB}(B)$, and $MSE^{EB}(B)$, for $B \in \{2\times 2, 4\times 4, \ldots 40\times 40\}$, centered on the center of the study domain. They are obtained empirically as follows. When using estimator A, the empirical MSE is:

$$\hat{M}SE^{A}(B) \equiv \frac{1}{L}\sum_{l=1}^{L}\left(CF^{A[l]}(B) - CF^{truth[l]}(B)\right)^2. \tag{26}$$

Based on (26), we now consider the relative MSE (REM) of the OB estimator to the EB estimator:

$$REM_{OB:EB}(B) \equiv \hat{M}SE^{OB}(B)/\hat{M}SE^{EB}(B); \tag{27}$$

a plot showing $REM_{OB:EB}(B)$ as the spatial support B increases is given in Figure 10. From the plot, we see that $REM_{OB:EB}(B)$ is generally close to 1 for the models with three spatial resolutions, for any spatial support B; in fact, up to the first decimal place, $REM_{OB:EB}(B) = 1$ for these models. For the model with four spatial resolutions we see that the $REM_{OB:EB}(B)$ is close to 1 for small spatial supports B and, as B increases, $REM_{OB:EB}(B)$ takes values around 0.9, indicating that EB is slightly less efficient at larger supports. Recall that in practice OB cannot be computed, but its empirical version EB can be.

Since the OB and the EB estimators in the simulation are unbiased, we can relate $REM_{OB:EB}(B)$ to the length of prediction intervals. When $REM_{OB:EB}(B) = 0.9$, it means that the length of the EB prediction interval is $1/\sqrt{0.9} = 1.05$ times the length of the OB prediction intervals. We also looked at histograms of prediction errors for the OB and the EB estimators, and we looked at scatter plots of OB estimates on the x-axis and EB estimates on the y-axis. All our analyses from the simulation study indicated that the EB estimate is a stable, near-optimal estimate of $CF(\cdot)$.

—— Figure 10 here. ——

We move on to the REM of the EB estimator with respect to the PI estimator, which is:

$$REM_{EB:PI}(B) \equiv \hat{M}SE^{EB}(B)/\hat{M}SE^{PI}(B); \tag{28}$$

a plot showing $REM_{EB:PI}(B)$ as the spatial support B increases is shown in Figure 11.

—— Figure 11 here.——

From Figure 11, we see that the EB estimator has much smaller MSE than the PI estimator, with $REM_{EB:PI}(B)$ decreasing as B increases. That is, prediction errors for the PI estimator are generally bigger than those for the corresponding EB estimator and, as the support size increases, the performance of the PI estimator, relative to EB estimator, worsens. Comparing the effect of the three levels of the resolution factor, one can see from Figure 11 that the performance of the EB estimator relative to the PI estimator is best for the model with four spatial resolutions, and that when we switch to a model with three spatial resolutions the performance deteriorates. Between the two spatial models at three resolutions, we see that $REM_{EB:PI}(B)$ is generally larger for the spatial model based on the three finer resolutions.

To conclude the simulation study, we note that the new EB estimator of $CF(\cdot)$ proposed in this article, namely $CF^{EB}(\cdot)$ given by (21), has much better predictive properties than the straightforward PI estimator that is currently in use within the operational MODIS algorithm package. Since the PI estimator does not use the spatial dependence in the data, its performance relative to the EB estimator deteriorates as the spatial dependence gets stronger.

6.4 Analysis of MODIS Cloud Data at Coarser Resolutions : Cloud Fraction at 5 km$\times$5 km Resolution

In this section, we estimate cloud fraction at a 5 km$\times$5 km spatial resolution from the 1 km$\times$1 km MODIS cloud mask. The noisy data $\mathbf{Q}_O$ are shown in the left panel of Figure 8, and we wish to upscale them to a coarser resolution. The 5 km$\times$5 km spatial support is of interest to geophysicists because the release of several Level-2 remote sensing products on 5 km$\times$5 km pixels depends on estimating cloud contamination of those pixels. For example, cloud-top pressure is computed on a 5 km$\times$5 km pixel if the estimated cloud fraction for that pixel is ≥ 0.16 (e.g., Pincus et al., 2012).

From the hierarchical spatial statistical model described in Section 2, and the MCMC samples from the predictive distribution, $[process|data, EM\text{-}estimated\ parameters]$, we can obtain the

predictive distribution, $[CF(B)|\mathbf{Q}_O, \hat{\boldsymbol{\theta}}_{EM}]$, where $B \subset D$ is a 5 km$\times$5 km pixel. Recall that

$$CF^{EB}(B) = E(CF(B)|\mathbf{Q}_O, \hat{\boldsymbol{\theta}}_{EM}),$$

which is obtained from the MCMC samples from the predictive distribution. The MCMC samples also enable us to quantify uncertainty associated with the predicted cloud fraction. In Figure 12, we show the PI estimator, the EB estimator, and the pixelwise difference between the PI and the EB estimator; we also show the standard deviation and the pixelwise 2.5 and 97.5 percentiles of the predictive distribution, $[CF(B)|\mathbf{Q}_O, \hat{\boldsymbol{\theta}}_{EM}]$. Further, the pixelwise 2.5 and 97.5 percentiles are the lower and upper endpoints of a pixelwise 95% prediction interval for $CF(B)$.

—— Figure 12 here. ——

Consider the example given above, where the geophysical question of interest is whether or not the cloud fraction for a 5 km$\times$5 km pixel is ≥ 0.16; if it is, cloud-top pressure is computed (Pincus et al., 2012). We use the EB estimator to answer this based on the data $\mathbf{Q}_O$. Figure 13 shows the 5 km$\times$5 km pixels where $CF^{EB}(B) \geq 0.16$ (marked light gray) and those where $CF^{EB}(B) < 0.16$ (marked as blue "+").

—— Figure 13 here. ——

7 Discussion and Conclusions

In this article, we give a hierarchical spatial statistical modeling approach for analyzing a very-large remote sensing dataset (2.75 million pixel values) on clouds from NASA's MODIS instrument. However, use of the reduced-rank SRE model to capture the spatial variability of the latent clear-sky-probability process allows for fast computations.

We have taken an empirical hierarchical modeling (EHM) approach, where the unknown model parameters are estimated using an EM algorithm. Alternatively, one could take a Bayesian hierarchical modeling (BHM) approach, where a prior distribution is put on the parameters. Kang and

Cressie (2011) developed the "Givens angle prior" for **K**, which could be adapted for analyzing the MODIS cloud data using a BHM. While the prediction intervals computed using the EHM approach tend to be liberal when compared to those using a BHM approach, EHM is an order of magnitude faster (e.g., Sengupta and Cressie, 2013b).

Within the hierarchical-statistical-modeling framework, we used the SRE model to define an underlying Gaussian field. Other than computational speed-ups, these models do not rely on specifying a spatial-weights matrix, and no assumptions of homogeneity, stationarity, or isotropy are needed. The hierarchical-statistical-modeling framework considered here enables us to make inferences on cloud fractions at resolutions coarser than 1 km$\times$1 km pixels. In some applications where researchers want to study local weather phenomena, interest lies in predicting the clear-sky probabilities (or equivalently, the probabilities of cloudiness) at resolutions finer than the 1 km$\times$1 km pixels. Obtaining high-resolution information from relatively coarse-resolution data is called downscaling.

When interest lies in downscaling, one can proceed by defining the *process models* for $W(\cdot)$ and $Y(\cdot)$ at the finest resolution at which we are interested in doing inference. Then, the *data model* will be specified at the (coarser) resolutions at which we have data, conditional on $W(\cdot)$ (process 1) and $Y(\cdot)$ (process 2) at the finest resolution. We can write:

$$\begin{aligned}
&[\textit{process}\ 1,\ \textit{process}\ 2|\textit{data},\ \textit{parameters}] \\
&\quad \propto [\textit{data}|\textit{process}\ 1,\ \textit{process}\ 2,\ \textit{parameters}] \times [\textit{process}\ 1|\textit{process}\ 2,\ \textit{parameters}] \\
&\quad\quad \times [\textit{process}\ 2|\textit{parameters}], \qquad (29)
\end{aligned}$$

where the data are at a coarser resolution than the processes. Hence, to carry out inference on $W(\cdot)$ (i.e., process 1), we could simulate from the predictive distribution,

$$[\textit{process}\ 1,\ \textit{process}\ 2|\textit{data},\ \textit{parameters}],$$

and use the simulated values of process 1 to obtain the optimal predictor of the cloud process,

$$E(I(W(\mathbf{s}) = 0)|data,\ parameters) = Pr(W(\mathbf{s}) = 0|data,\ parameters),$$

where $\mathbf{s}$ is a location at the finest resolution.

To our knowledge, this article represents the first attempt to develop a hierarchical spatial statistical model for clouds at such a fine resolution. The model developed here could be extended to a spatio-temporal setting that might be used to improve a climate model's subgrid-scale physical parameterization. Another extension would be to develop a data-fusion methodology, such as was done by Nguyen et al. (2012). For example, one might fuse water-vapor data from NASA's AIRS instrument with cloud data from MODIS; both are on the Terra satellite.

Acknowledgments

This research was supported by NASA's Earth Science Technology Office through its Advanced Information Systems Technology Program. It was also partially supported by the Orbiting Carbon Observatory Project at Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA. We are grateful to Amy Braverman, Mathias Schreier, and Robert Pincus for their generous contributions to this research.

References

Ackerman, S. A., Strabala, K. I., Menzel, W. P., Frey, R. A., Moeller, C. C., and Gumley, L. E. (1998). "Discriminating clear sky from clouds with MODIS." *Journal of Geophysical Research - Atmospheres*, 103, 32141–32157.

— (2010). "Discriminating clear sky from clouds with MODIS algorithm theoretical basis document (MOD35)." Version 6.1. Website: `http://modis.gsfc.nasa.gov/data/atbd/atbd_mod06.pdf`.

Aldworth, J. and Cressie, N. (1999). "Sampling designs and prediction methods for Gaussian spatial processes." In *Multivariate Analysis, Designs of Experiments, and Survey Sampling*, ed. S. Ghosh, 1–54. New York, NY: Markel Dekker, Inc.

Banerjee, S., Gelfand, A. E., Finley, A. O., and Sang, H. (2008). "Gaussian predictive process models for large spatial data sets." *Journal of the Royal Statistical Society, Series B*, 70, 825–848.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. 2nd ed. New York, NY: Springer-Verlag.

Bradley, J. R., Cressie, N., and Shi, T. (2011). "Selection of rank and basis functions in the Spatial Random Effects model." In *Proceedings of the 2011 Joint Statistical Meetings*, 3393–3406. Alexandria, VA: American Statistical Association.

Chib, S. and Greenberg, E. (1995). "Understanding the Metropolis-Hastings algorithm." *The American Statistician*, 49, 327–335.

— (1996). "Markov chain Monte Carlo simulation methods in econometrics." *Econometric Theory*, 12, 409–431.

Cressie, N. (2006). "Block kriging for lognormal spatial processes." *Mathematical Geology*, 38, 413–443.

Cressie, N. and Johannesson, G. (2006). "Spatial prediction for massive data sets." In *Australian Academy of Science Elizabeth and Frederick White Conference*, 1–11. Canberra, Australia: Australian Academy of Science.

— (2008). "Fixed rank kriging for very large spatial data sets." *Journal of the Royal Statistical Society, Series B*, 70, 209–226.

Cressie, N. and Kang, E. (2010). "High-resolution digital soil mapping: Kriging for very large datasets." In *Proximal Soil Sensing*, eds. R. A. Viscarra-Rossel, A. B. McBratney, and B. Minasny, vol. 1 of *Progress in Soil Science*, 49–63. Dordrecht, NL: Springer.

Cressie, N. and Wikle, C. K. (2011). *Statistics for Spatio-Temporal Data*. Hoboken, NJ: Wiley.

Dempster, A. P., Laird, N., and Rubin, D. (1977). "Maximum likelihood from incomplete data via the EM algorithm." *Journal of the Royal Statistical Society, Series B*, 39, 1–38.

Di Girolamo, L. and Davies, R. (1997). "Cloud fraction errors caused by finite resolution measurements." *Journal of Geophysical Research - Atmospheres*, 102, 1739–1756.

Diggle, P. J., Tawn, J. A., and Moyeed, R. A. (1998). "Model-based geostatistics." *Journal of the Royal Statistical Society, Series C*, 47, 299–350.

Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap*. Boca Raton, FL: Chapman & Hall/CRC.

Hughes, J. and Haran, M. (2012). "Dimension reduction and alleviation of confounding for spatial generalized linear mixed models." *Journal of the Royal Statistical Society, Series B*, forthcoming.

Kang, E. L. and Cressie, N. (2011). "Bayesian inference for the Spatial Random Effects model." *Journal of the American Statistical Association*, 106, 972–983.

Katzfuss, M. and Cressie, N. (2009). "Maximum likelihood estimation of covariance parameters in the spatial-random-effects model." In *Proceedings of the 2009 Joint Statistical Meetings*, 3378–3390. Alexandria, VA: American Statistical Association.

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." *Journal of the Royal Statistical Society, Series B*, 73, 423–498.

Lopes, H. F., Gamerman, D., and Salazar, E. (2011). "Generalized spatial dynamic factor models." *Computational Statistics and Data Analysis*, 55, 1319 – 1330.

Lopes, H. F., Salazar, E., and Gamerman, D. (2008). "Spatial dynamic factor analysis." *Bayesian Analysis*, 3, 759–792.

McLachlan, G. J. and Krishnan, T. (2008). *The EM Algorithm and Extensions*. 2nd ed. New York, NY: Wiley-Interscience.

Nguyen, H., Cressie, N., and Braverman, A. (2012). "Spatial statistical data fusion for remote sensing applications." *Journal of American Statistical Association*, 107, 1004–1018.

Pincus, R., Platnick, S., Ackerman, S. A., Hemler, R. S., and Hofmann, R. J. P. (2012). "Reconciling simulated and observed views of clouds: MODIS, ISCCP, and the limits of instrument simulators." *Journal of Climate*, 25, 4699–4720.

Platnick, S., King, M., Ackerman, S., Menzel, W., Baum, B., Riedi, J., and Frey, R. (2003). "The MODIS cloud products: algorithms and examples from Terra." *IEEE Transactions on Geoscience and Remote Sensing*, 41, 459 – 473.

Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods*. New York, NY: Springer.

Roberts, G. O., Gelman, A., and Gilks, W. R. (1997). "Weak convergence and optimal scaling of random walk Metropolis algorithms." *Annals of Applied Probability*, 7, 110–120.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*. London, UK: Chapman & Hall/CRC.

Sengupta, A. (2012). "Empirical Hierarchical Modeling and Predictive Inference for Big, Spatial, Discrete, and Continuous Data." PhD Thesis, The Ohio State University, Columbus, OH.

Sengupta, A. and Cressie, N. (2013a). "Empirical hierarchical modeling for count data using the Spatial Random Effects model." *Spatial Economic Analysis*, 8, 389–418.

— (2013b). "Hierarchical statistical modeling of big spatial datasets using the exponential family of distributions." *Spatial Statistics*, 4, 14–44.

Sengupta, A., Cressie, N., Frey, R., and Kahn, B. H. (2012). "Statistical modeling of MODIS cloud data using the Spatial Random Effects model." In *Proceedings of the 2012 Joint Statistical Meetings*, 3111–3123. Alexandria, VA: American Statistical Association.

Shi, T. and Cressie, N. (2007). "Global statistical analysis of MISR aerosol data: a massive data product from NASA's Terra satellite." *Environmetrics*, 18, 665–680.

Simpson, D., Lindgren, F., and Rue, H. (2012). "Think continuous: Markovian Gaussian models in spatial statistics." *Spatial Statistics*, 1, 16 – 29.

Stein, M. L. (2008). "A modeling approach for large spatial datasets." *Journal of the Korean Statistical Society*, 37, 3 – 10.

Tierney, L. (1994). "Markov chains for exploring posterior distributions." *Annals of Statistics*, 22, 1701–1728.

Wikle, C. K. (2010). "Low-rank representations for spatial processes." In *Handbook of Spatial Statistics*, eds. A. E. Gelfand, P. J. Diggle, M. Fuentes, and P. Guttorp, 107–118. Boca Raton, FL: Chapman and Hall/CRC.

Wikle, C. K. and Cressie, N. (1999). "A dimension-reduced approach to space-time Kalman filtering." *Biometrika*, 86, 815–829.

Wikle, C. K., Milliff, R. F., Nychka, D., and Berliner, L. M. (2001). "Spatiotemporal hierarchical Bayesian modeling: Tropical ocean surface winds." *Journal of the American Statistical Association*, 96, 382–397.

Zhao, G. and Di Girolamo, L. (2006). "Cloud fraction errors for trade wind cumuli from EOS-Terra instruments." *Geophysical Research Letters*, 33, L20802 (5 pp.).

Appendix

A MCMC Algorithm used in Section 5

The joint distribution, $[\mathbf{Q}_O, \mathbf{W}, \boldsymbol{\eta}, \boldsymbol{\xi} | \boldsymbol{\beta}, \mathbf{K}, \sigma_\xi^2]$, can be written as:

$$[\mathbf{Q}_O, \mathbf{W}, \boldsymbol{\eta}, \boldsymbol{\xi} | \boldsymbol{\beta}, \mathbf{K}, \sigma_\xi^2] \equiv [\mathbf{Q}_O | \mathbf{W}, \boldsymbol{\eta}, \boldsymbol{\xi}, \boldsymbol{\beta}] \times [\mathbf{W} | \boldsymbol{\eta}, \boldsymbol{\xi}, \boldsymbol{\beta}] \times [\boldsymbol{\eta} | \mathbf{K}] \times [\boldsymbol{\xi} | \sigma_\xi^2]. \tag{30}$$

Let "$[\mathbf{A}|\mathbf{B}, \cdot]$" denote the full conditional distribution of the unknown **A** given **B** and all other unknowns (and the data). The Gibbs sampler uses the following steps to generate samples from the predictive distribution, $[\mathbf{W}, \boldsymbol{\eta}, \boldsymbol{\xi} | \mathbf{Q}_O, \boldsymbol{\beta}, \mathbf{K}, \sigma_\xi^2]$.

1. At $t = 0$, select starting values $\mathbf{W}^{[0]}$, $\boldsymbol{\eta}^{[0]}$, and $\boldsymbol{\xi}^{[0]}$.

2. t=t+1; simulate successively from the full conditionals, $[\mathbf{W}^{[t+1]} | \boldsymbol{\eta}^{[t]}, \boldsymbol{\xi}^{[t]}, \cdot]$, $[\boldsymbol{\eta}^{[t+1]} | \mathbf{W}^{[t+1]}, \boldsymbol{\xi}^{[t]}, \cdot]$, and $[\boldsymbol{\xi}^{[t+1]} | \mathbf{W}^{[t+1]}, \boldsymbol{\eta}^{[t+1]}, \cdot]$.

3. Repeat step 2 to generate as many samples as needed.

4. Discard an initial number of samples as "burn-in."

The full conditional, $[\mathbf{W}^{[t+1]} | \boldsymbol{\eta}^{[t]}, \boldsymbol{\xi}^{[t]}, \cdot]$, is available in closed form, and it is straightforward to simulate from this distribution. The full conditionals, $[\boldsymbol{\eta}^{[t+1]} | \mathbf{W}^{[t+1]}, \boldsymbol{\xi}^{[t]}, \cdot]$ and $[\boldsymbol{\xi}^{[t+1]} | \mathbf{W}^{[t+1]}, \boldsymbol{\eta}^{[t+1]}, \cdot]$, are not available in closed form, so we use a Metropolis step within the Gibbs sampler. A generic version of this Metropolis algorithm is now presented.

Suppose **a** is the random variable (or a block of random variables) that are being updated, and $\mathbf{a}_0$ is the most recently sampled value. We follow the steps below to obtain a new sample of **a**:

1. Draw a trial value $\mathbf{a}_1$ from a proposal density, $g(\mathbf{a})$.

2. Generate U_1 uniformly on $(0, 1)$.

3. Compute the joint density of **a** and all other unknowns, and obtain $l(\mathbf{a}_0, rest)$ and $l(\mathbf{a}_1, rest)$, where "*rest*" denotes all the other unknowns fixed at their most recently sampled value.

4. If $U_1 < \min\left\{\frac{l(\mathbf{a}_1,rest)g(\mathbf{a}_0)}{l(\mathbf{a}_0,rest)g(\mathbf{a}_1)}\right\}$, accept the new $\mathbf{a}_1$ and keep it for the most current iteration; otherwise, the value $\mathbf{a}_0$ is retained.

When sampling from $[\boldsymbol{\eta}^{[t+1]}|\boldsymbol{\xi}^{[t]},\cdot]$, we updated $\boldsymbol{\eta}$ as a block. For the proposal, we used a mixture distribution (see Tierney, 1994). Specifically, we used a mixture of a random-walk proposal (see Robert and Casella, 2004, Section 7.5) and a tailored multivariate normal proposal (see Chib and Greenberg, 1995, 1996). Let $\boldsymbol{\eta}_{mode}$ denote the posterior mode of $[\boldsymbol{\eta}|\boldsymbol{\xi},\cdot]$ and $\boldsymbol{\Sigma}_\eta$ denote the inverse of the negative Hessian matrix, evaluated at the mode. The mode can be computed using a Newton-Raphson method (see Robert and Casella, 2004, Section 1.4). The tailored proposal for $\boldsymbol{\eta}$ was then taken as $\text{Gau}(\boldsymbol{\eta}_{mode}, c_{1,\eta}\boldsymbol{\Sigma}_\eta)$, where $c_{1,\eta}$ is a tuning parameter that is adjusted to control the acceptance rate. The random walk proposal was taken to be $\text{Gau}(\boldsymbol{\eta}^{[t]}, c_{2,\eta}\boldsymbol{\Sigma}_\eta)$, where again $c_{2,\eta}$ is a tuning parameter. The mixing probability could also be tuned to achieve the desired acceptance rate. We worked with a mixing probability of 0.5 and selected $c_{1,\eta} = 1$ and $c_{2,\eta} = 2.3$. With these values for the tuning parameters, we achieved an acceptance rate of 22.1% for $\boldsymbol{\eta}$, which is close to optimal (see Roberts et al., 1997).

When sampling from $[\boldsymbol{\xi}^{[t+1]}|\boldsymbol{\eta}^{[t+1]},\cdot]$, we updated $\boldsymbol{\xi}$ elementwise. Here we worked with the tailored proposal. We computed the mode of $[\boldsymbol{\xi}|\boldsymbol{\eta},\cdot]$, which we denote as $\boldsymbol{\xi}_{mode}$, and the inverse of the negative Hessian matrix evaluated at the mode, which we denote as $\boldsymbol{\Sigma}_\xi$. For updating the i-th element of $\boldsymbol{\xi}$, we used the proposal: $\text{Gau}\left((\boldsymbol{\xi}_{mode})_i, c_\xi\left(\boldsymbol{\Sigma}_\xi\right)_{i,i}\right)$, where $(\boldsymbol{\xi}_{mode})_i$ denotes the i-th element of $\boldsymbol{\xi}_{mode}$, and $\left(\boldsymbol{\Sigma}_\xi\right)_{i,i}$ denotes the (i,i)-th element of the matrix $\boldsymbol{\Sigma}_\xi$. We fixed $c_\xi = 8$ and achieved an acceptance rate between 40.53% and 46.13% for the different elements of $\boldsymbol{\xi}$.

Figures

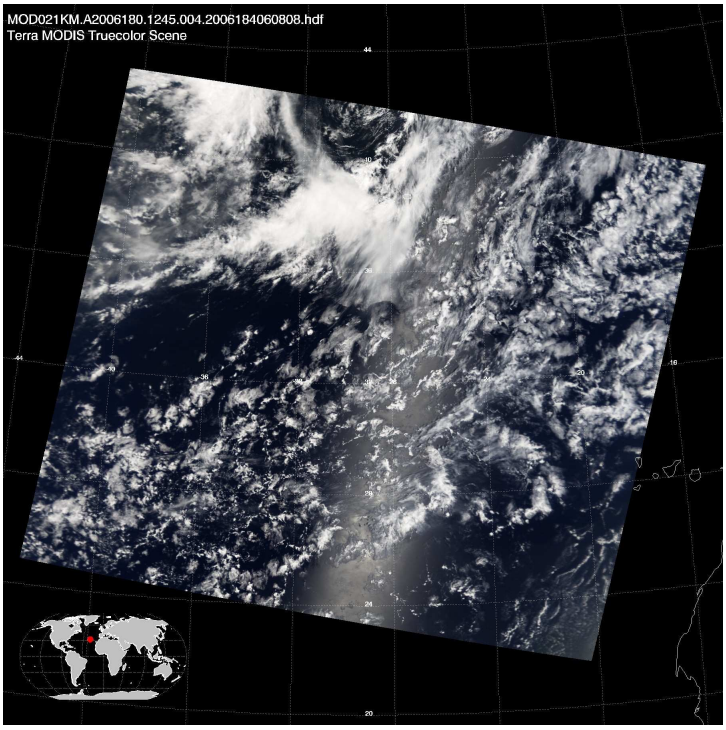

Figure 1: Image of a granule obtained by the MODIS instrument on board NASA's Terra satellite (June 29, 2006, 12:45 UTC). The inset shows the location of the granule on a world map. (Source: `modis-atmos.gsfc.nasa.gov`)

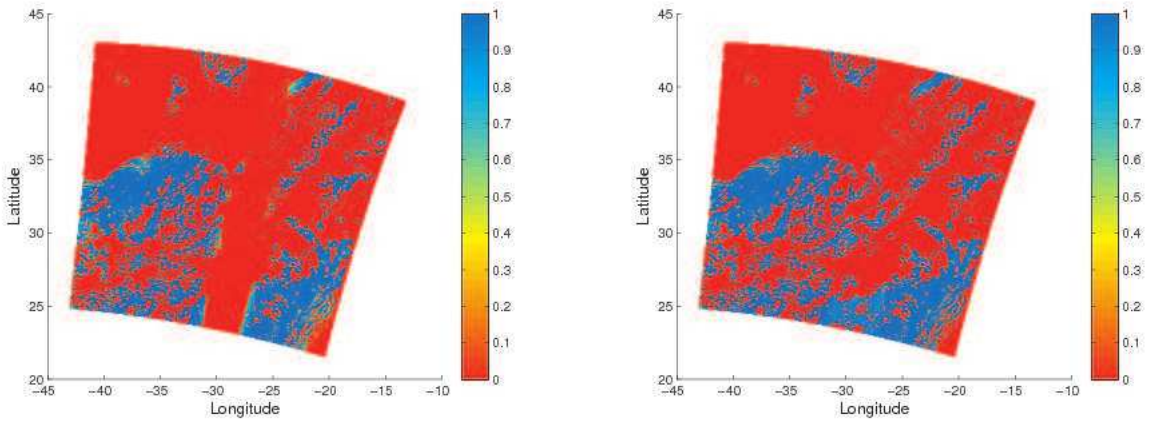

Figure 2: Initial Q-values (left panel) and final Q-values (right panel) corresponding to the granule shown in Figure 1.

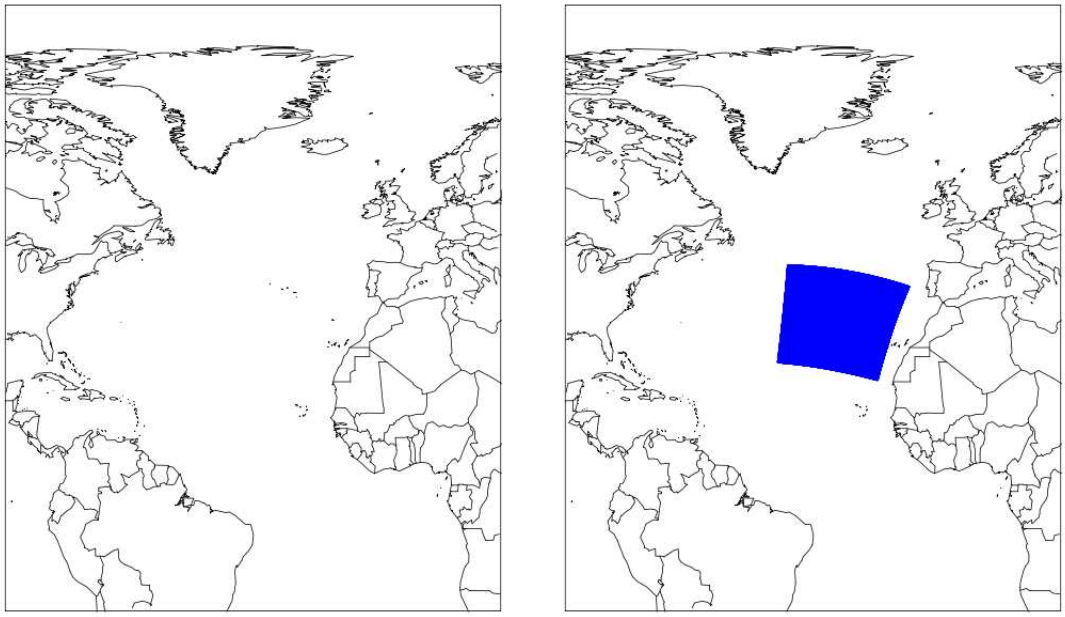

Figure 3: Geographical location of the data granule on June 29, 2006, 12:45 UTC

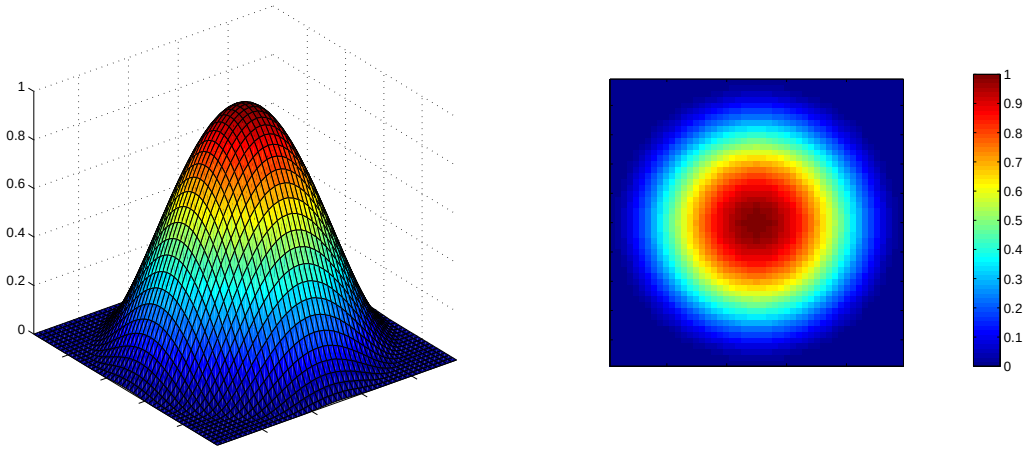

Figure 4: A bisquare function defined on $\Re^2$, illustrated as a 3-D plot (left panel) and as an image plot (right panel).

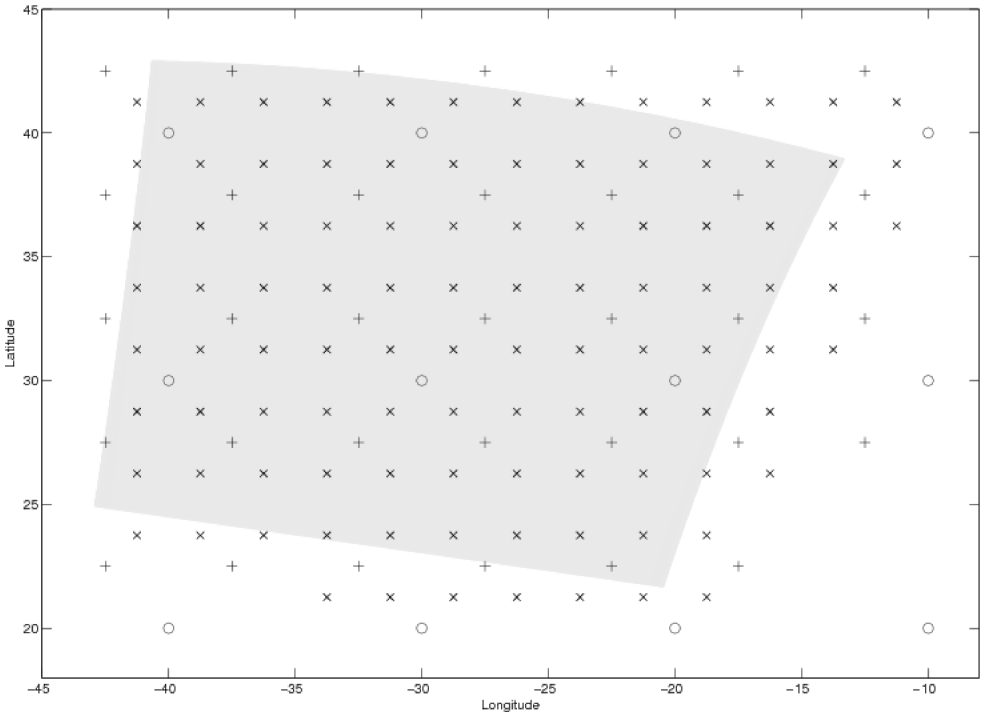

Figure 5: Centers of the basis function; ’o’, ’+’, and ’x’ are use to distinguish the three scales of resolution.

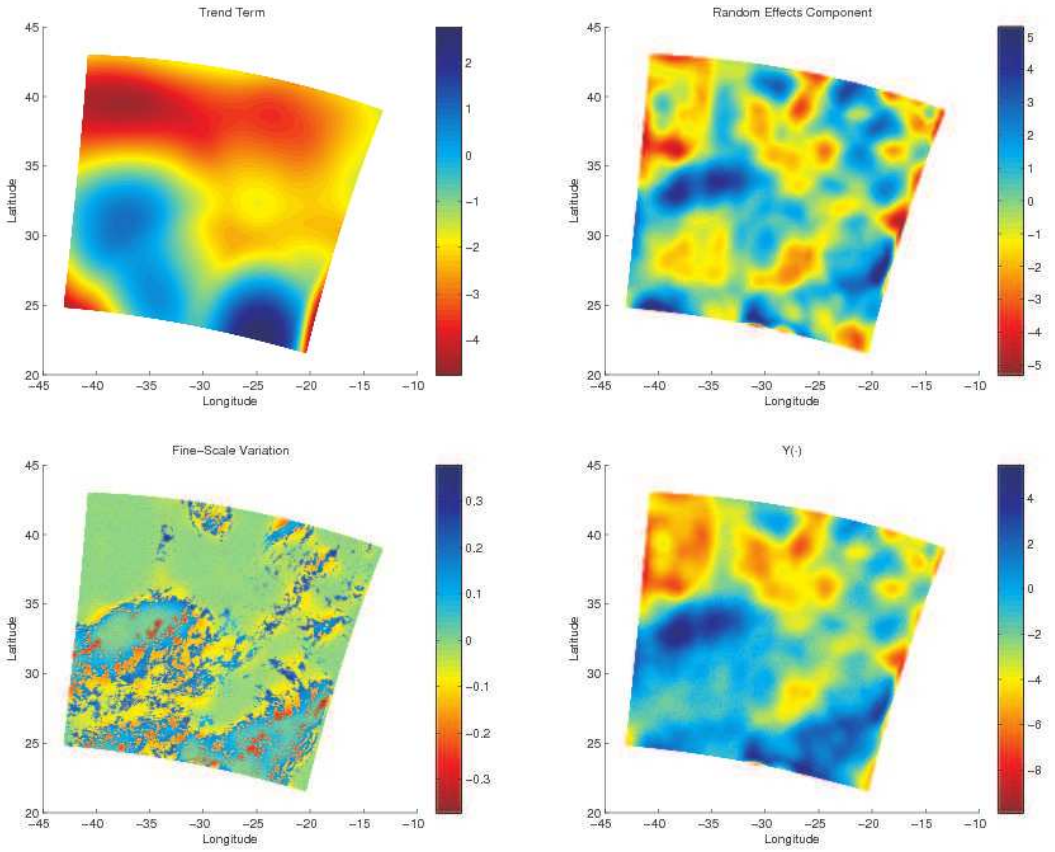

Figure 6: Maps showing predictions of the components of variation: the trend, $\mathbf{X}(\cdot)^\top \hat{\boldsymbol{\beta}}_{EM}$ (top-left panel), the mean of the predictive distribution of the random-effects component, $E[\mathbf{S}(\cdot)^\top \boldsymbol{\eta} | \mathbf{Q}, \hat{\boldsymbol{\theta}}_{EM}]$ (top-right panel), and the mean of the predictive distribution of the component due to fine-scale variation, $E[\xi(\cdot)|\mathbf{Q}, \hat{\boldsymbol{\theta}}_{EM}]$ (bottom-left panel). The mean of the predictive distribution of the hidden process, $E[Y(\cdot)|\mathbf{Q}_O, \hat{\boldsymbol{\theta}}_{EM}]$ (bottom-right panel), is equal to the sum of the previous three components.

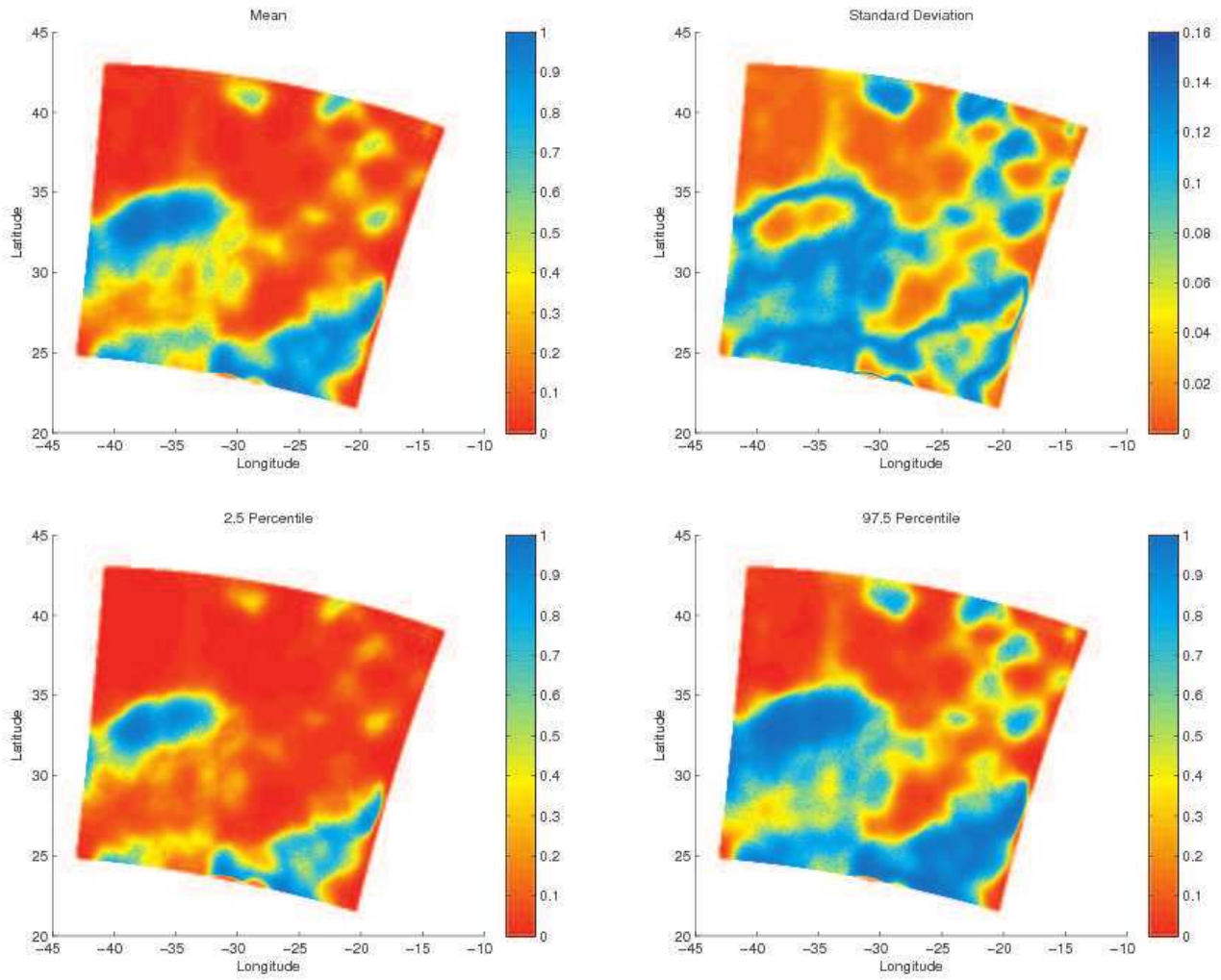

Figure 7: Maps showing the predictive mean (top-left panel), the pixelwise predictive standard-deviation (top-right panel), the pixelwise 2.5 percentile (bottom-left panel) and the pixelwise 97.5 percentile (bottom-right panel) for the predictive distribution of the clear-sky-probability process.

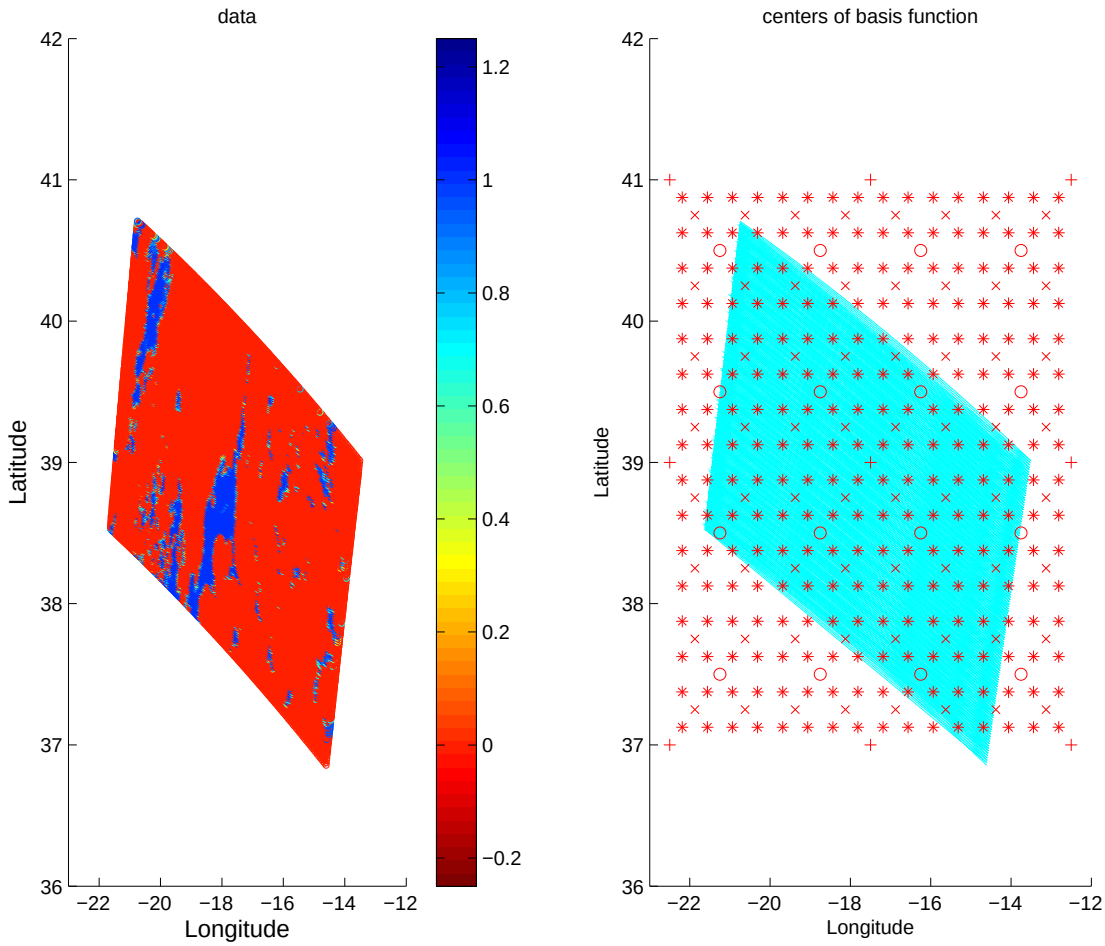

Figure 8: Data for the sub-granule that we use to illustrate change-of-support (left panel). Spatial region of interest, D, and centers of basis function at the four scales of resolution considered in the simulation study (right panel).

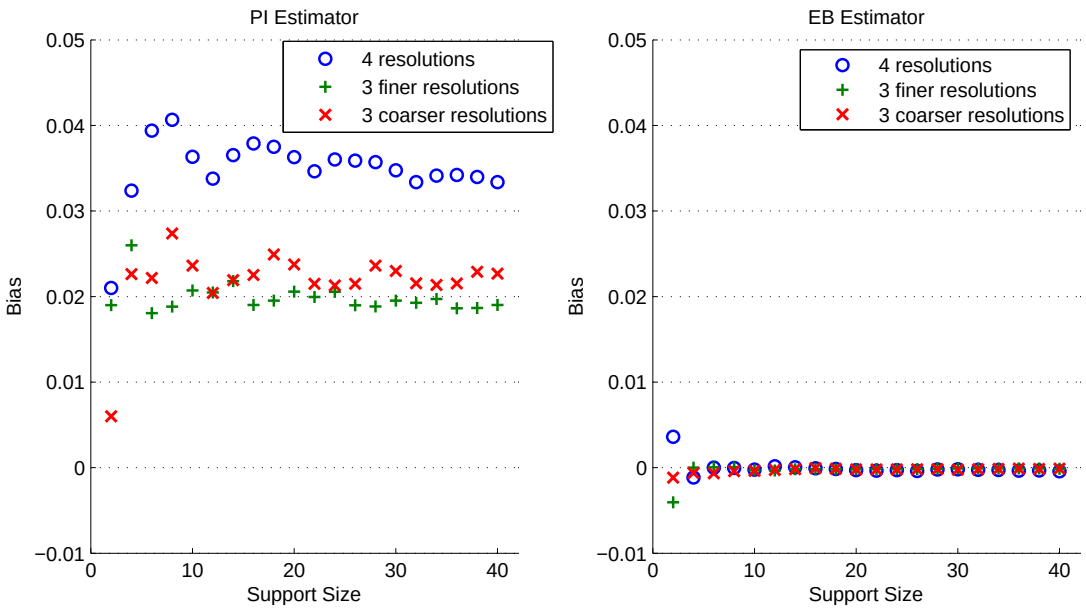

Figure 9: Plots showing the estimation bias. The left panel shows bias for the plug-in (PI) estimator, and the right panel shows bias for the empirical Bayes (EB) estimator. The different plot symbols (and colors) correspond to the three levels of the resolution factor.

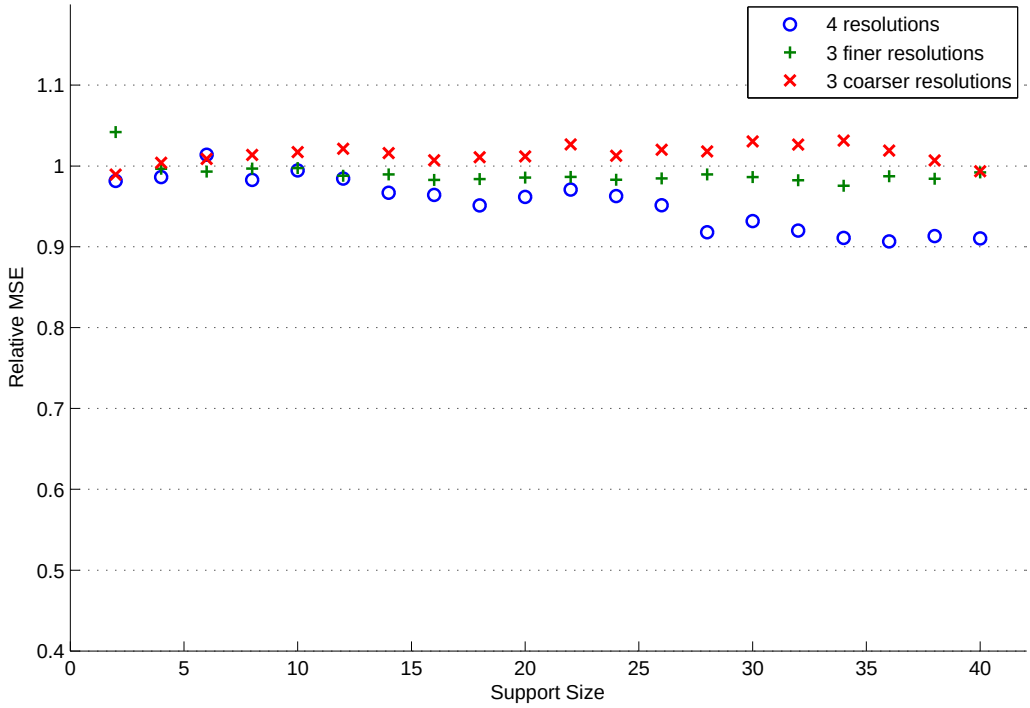

Figure 10: Plot showing the relative MSE of the OB estimator relative to the EB estimator. The different plot symbols (and colors) correspond to the three levels of the resolution factor.

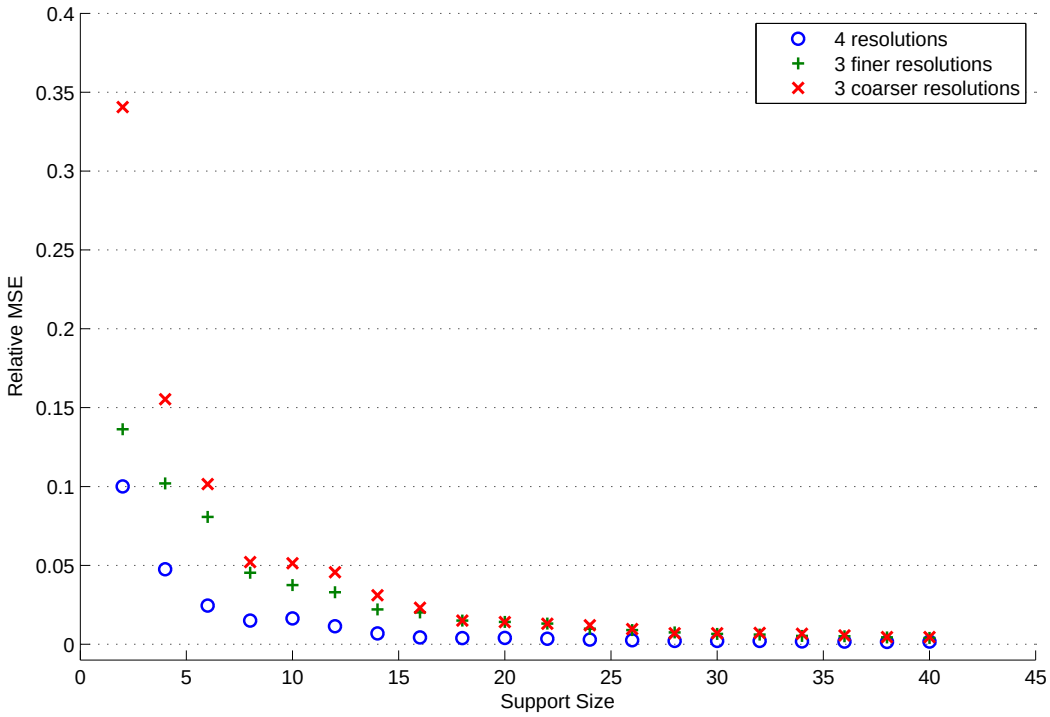

Figure 11: Plot showing the relative MSE of the EB estimator relative to the PI estimator. The different plot symbols (and colors) correspond to the three levels of the resolution factor.

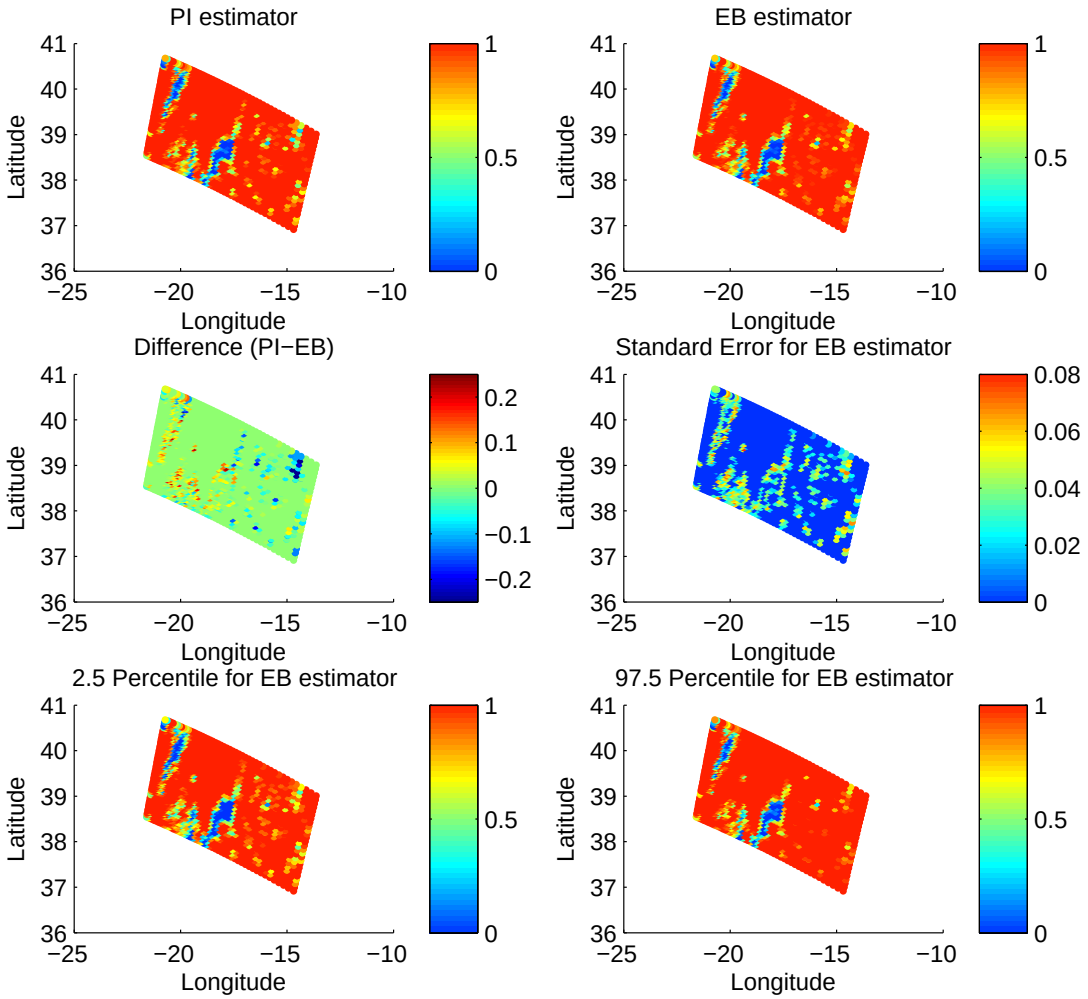

Figure 12: Maps showing the PI estimator (top-left panel), the EB estimator (top-right panel), the pixelwise difference (PI-EB) between the two estimators (middle-left panel), the standard deviation from the (empirical) predictive distribution (middle-right panel), the pixelwise 2.5 percentile from the (empirical) predictive distribution (bottom-left panel), and the pixelwise 97.5 percentile from the (empirical) predictive distribution (bottom-right panel).

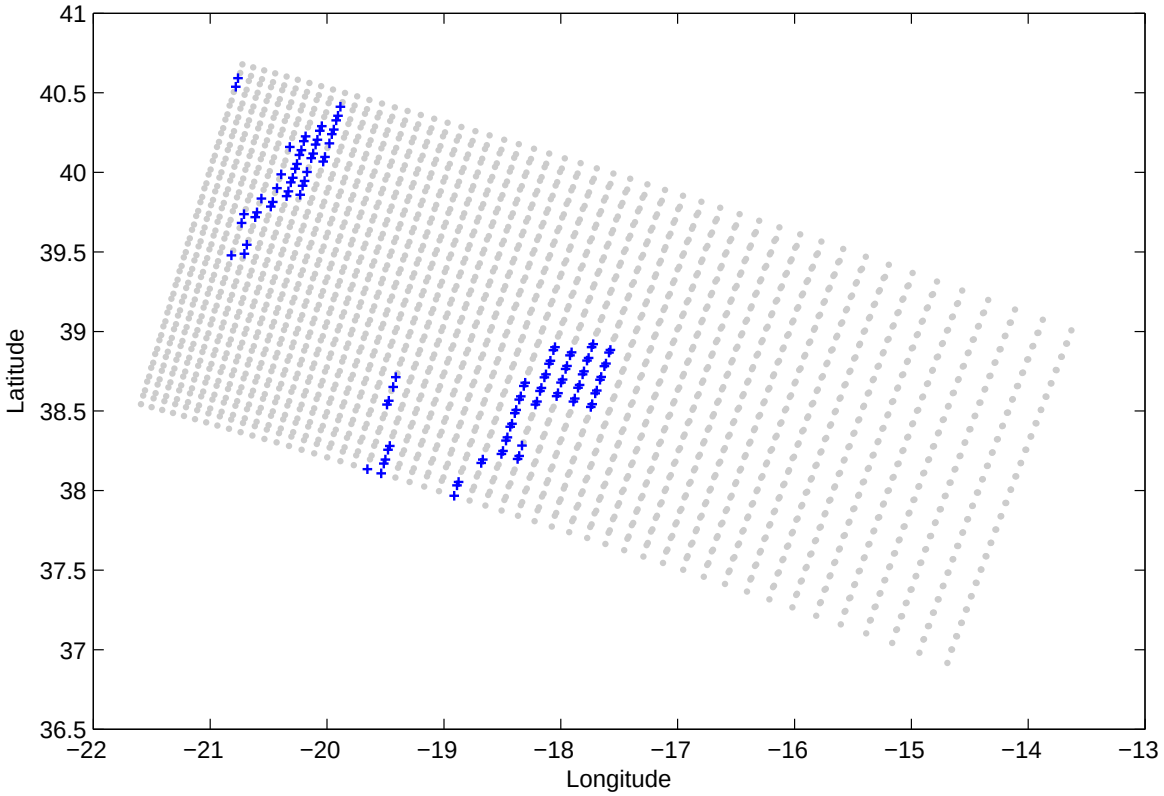

Figure 13: Plot showing the 5 km$\times$5 km pixels where $CF^{OB}(\cdot) \geq 0.16$ (marked light gray) and the pixels where $CF^{OB}(\cdot) < 0.16$ (marked as blue "+").