Stat Comput (2015) 25:389–405
DOI 10.1007/s11222-013-9439-8

An adaptive spatial model for precipitation data from multiple satellites over large regions

Avishek Chakraborty · Swarup De · Kenneth P. Bowman · Huiyan Sang · Marc G. Genton · Bani K. Mallick

Received: 2 December 2012 / Accepted: 20 November 2013 / Published online: 4 December 2013

Abstract Satellite measurements have of late become an important source of information for climate features such as precipitation due to their near-global coverage. In this article, we look at a precipitation dataset during a 3-hour window over tropical South America that has information from two satellites. We develop a flexible hierarchical model to combine instantaneous rainrate measurements from those satellites while accounting for their potential heterogeneity. Conceptually, we envision an underlying precipitation surface that influences the observed rain as well as absence of it. The surface is specified using a mean function centered at a set of knot locations, to capture the local patterns in the rainrate, combined with a residual Gaussian process to account for global correlation across sites. To improve over the commonly used pre-fixed knot choices, an efficient reversible jump scheme is used to allow the number of such knots as well as the order and support of associated polynomial terms to be chosen adaptively. To facilitate computation over a large region, a reduced rank approximation for the parent Gaussian process is employed.

Keywords Large data computation · Nonstationary spatial model · Precipitation modeling · Predictive process · Random knots · Reversible jump Markov chain Monte Carlo · Satellite measurements

1 Introduction

Algorithms to estimate atmospheric parameters from satellite measurements of upwelling radiation (Lethbridge 1967) have become invaluable for investigating global weather and climate. Satellites are routinely used to observe temperature, humidity, clouds, precipitation, aerosols and atmospheric trace constituents. Applications of satellite data include weather forecasting, climate studies, ozone depletion, drought monitoring, crop forecasting and flood warning. Data obtained from satellites are attractive because of their potential for near-global coverage and their ability to generate measurements at high spatial and temporal resolution compared to ground-based or airborne sources. Over the ocean and in sparsely populated regions of the land surface where few ground-based measurements exist, satellite observations are essential (Simpson et al. 1988; Kidd 2001).

The research of Bani K. Mallick and Avishek Chakraborty was supported by National Science Foundation grant DMS 0914951. Research of Marc G. Genton was partially supported by NSF grants DMS-1007504 and DMS-1100492. The research in this article was also partially supported by Award No. KUSC1-016-04 made by King Abdullah University of Science and Technology (KAUST).

A. Chakraborty (✉) · H. Sang · B.K. Mallick
Department of Statistics, Texas A&M University, College Station, TX 77843-3143, USA
e-mail: avishekc@stat.tamu.edu

H. Sang
e-mail: huiyan@stat.tamu.edu

B.K. Mallick
e-mail: bmallick@stat.tamu.edu

S. De
SAS Research & Development (India) Pvt. Ltd, Pune 411013, India
e-mail: swarup.de@sas.com

K.P. Bowman
Department of Atmospheric Sciences, Texas A&M University, College Station, TX 77843-3150, USA
e-mail: k-bowman@tamu.edu

M.G. Genton
CEMSE Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia
e-mail: marc.genton@kaust.edu.sa

Precipitation is critically important in terms of economic and social impacts, but it is also one of the most difficult atmospheric phenomena to observe and model. The complex physical processes and large variability of precipitation pose significant scientific challenges in modeling the atmosphere. Rain gauges are the most common technology for measuring surface rainrates. Compared to parameters such as temperature, rainfall has very short space and time correlation length scales, so high resolution data are required to resolve spatial and temporal variations (Austin and Houze 1972; Rodríguez-Iturbe and Mejía 1974). With the exception of a few small, dense research networks, the spacing between rain gauges tends to be much larger than the typical spatial scales of precipitation systems (Felgate and Read 1975). In practical terms, gauge networks do not resolve the variability of precipitation systems, particularly on smaller scales. In addition, there is little rain gauge data over the oceans and over large expanses of many continents. In contrast, due to their near-global coverage, satellites have the potential to provide precipitation estimates with high spatial resolution where gauge observations or conventional earth-bound monitoring systems are unavailable. The principal limitation of observations from low-Earth-orbiting satellites is temporal sampling. A single satellite will typically observe a given location on the Earth's surface only a few times per day. Although satellites can provide global coverage, satellite sampling and measurement errors can be substantial, so effective methods are needed to validate and combine observations into consistent and useful estimates. A detailed discussion on sampling errors for satellite rainfall average can be found in Bell et al. (1990, 2001), Bell and Kundu (1996) and McConnell and North (1987).

Over the last few decades many satellite-based precipitation algorithms have been developed (Wilheit 1977; Xie and Arkin 1998; Ba and Gruber 2001; Huffman et al. 2002; Joyce et al. 2004; Negri et al. 2002; Sorooshian et al. 2000; Vicente et al. 1998; Weng et al. 2003). Satellite methods can be used to generate precipitation products at various spatial and temporal resolutions. Precipitation estimation from space is most often based on observations of infrared or microwave radiation. Infrared-based techniques can have relatively high spatial and temporal resolution. Infrared radiation cannot penetrate typical dense clouds, but it is possible to measure the altitude of cloud tops. Higher cloud tops indicate deeper storms, which tend to produce larger rainfall amounts. A statistical relationship between cloud-top height and surface rain gauge data can be used to estimate surface rainrates from satellite observations. Hence, the rainrate is related only indirectly to the observed quantity (cloud top height); so uncertainties are larger. Microwave radiation, on the other hand, can penetrate through clouds. Consequently, microwave methods are more closely tied to the relevant physical quantity (falling raindrops), but spatial and temporal coverage are typically lower. Microwave methods address these problems by merging precipitation observations from multiple satellites to yield global precipitation estimates at reasonably high spatial and temporal resolution (e.g. 0.25° and a few hours) as in the Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA) described by Huffman et al. (2007). The resolution of the TMPA grid (0.25°) is comparable to the resolution of current microwave instruments. Currently merged satellite estimates do not take into account the different statistical properties of the input data streams. These differences include variations in spatial and temporal resolution and in error characteristics. Multiple observations are usually combined in the simplest possible way by averaging the various instantaneous estimates to produce a "best estimate" of the mean rainfall rate over the selected interval (Huffman et al. 2007). Because statistical properties of rainfall are highly non-Gaussian and depend strongly on space and time scale, this averaging can significantly impact higher moments of the estimates (e.g. variances and covariances).

In this paper, our approach is to use a Bayesian hierarchical framework for combining the available observations from multiple satellites within a space-time volume. This enables us to develop the model in a way that can account for some specific characteristics of a precipitation dataset. Generally, precipitation patterns are highly localized and have fast time scales. As a result, at any given location, rainfall is absent most of the time. Hence any rainfall data, aggregated over a short time interval, is expected to have a high number of zeros. The use of mixture models with a degenerate mass at zero is a common approach to model zero-inflated data, e.g., a zero inflated Poisson (ZIP) model (Cohen 1991). Agarwal et al. (2002) discussed the use of Bayesian methods to analyze spatially correlated zero-inflated count data in the presence of covariate information. In the context of an ecological dataset on presence of plant species, Chakraborty et al. (2010) used a spatial probit model to address a large number of absences. In the current work, we introduce a locally varying bias term for that. Atmospheric convection, which drives the event of precipitation, involves turbulent interactions across a wide range of scales and is governed by fundamentally nonlinear equations of fluid dynamics. As a consequence, the information on whether it is raining at a given location or not is highly localized. The bias term we introduce here addresses this nonlinearity and local variation, unlike the linearity assumption in Fuentes et al. (2008). For geospatial datasets, an additive model with nonlinear covariate-dependence was also discussed in Kammann and Wand (2003).

For precipitation, when it rains, the probability distribution of instantaneous rainrate is non-Gaussian with long tail. Hence, a lognormal model is often adopted for nonzero precipitation measurements, see Lee and Zawadzki (2005) and

Fuentes et al. (2008). Alternative approaches also exist in the literature such as the truncated power transformation of Bardossy and Plate (1992) and the use of skew-elliptical distributions in Marchenko and Genton (2010). Our approach is motivated by Fuentes et al. (2008), where a zero-inflated log-Gaussian model has been used for rainfall, and, in a hierarchical framework, the distribution of no rainfall events was modeled to depend on the true rainfall intensity. However, the modeling of the rainfall process in this article significantly differs from the approach therein. To jointly model true rainfall intensities at adjacent locations, Fuentes et al. (2008) used a Gaussian Markov random field with correlation parameters that do *not* change across locations. However, in general, the amount of rainfall over a few hours is highly localized and nonstationary on a large domain. A number of works have focussed on developing nonstationary spatial covariance functions. Spatial deformations to model nonstationary spatial processes have been used by Sampson and Guttorp (1992), Schmidt and O'Hagan (2003) and Anderes and Stein (2008). On the other hand, kernel convolution and its variants have been applied in several papers to create nonstationary covariance functions as in Higdon (1998), Fuentes (2002) and Paciorek and Schervish (2006). Jun and Stein (2008) and Jun (2011) proposed a method of modeling nonstationary covariance functions on a sphere. Nonstationarity can also be introduced through covariates. Using spatially-varying regression coefficients for a point-referenced data provides a scope of detecting subregional variation in the response-predictor relationship, see Gelfand et al. (2003). Specifically, in the context of analyzing a zero-inflated dataset like ours, Finley et al. (2011) proposed a hierarchical model where multivariate spatial process priors were used for two different sets of regression coefficients—one for controlling the abundance of zeros and the other for the nonzero observations. Here, we also work with a covariate-dependent nonstationary model but opt for an adaptive specification, as in Friedman (1991) and Denison et al. (1998). It depends on finding a set of *knot* locations in the predictor space and developing a local polynomial for each one of them. This approach is relatively simpler to interpret, estimate and, importantly, model non-Gaussian patterns and input interactions in the response surface. The flexibility of this specification lies in the fact that the functions can be constructed adaptively, i.e., the order and support of the local polynomials and even the number of knots is decided by the pattern of the data during model-fitting. When spatial covariates are available, the model can select the important predictors and/or interactions, eliminating the need to pre-specify the form of dependence. To modify the number of polynomial terms in the function, an efficient reversible jump Markov chain Monte Carlo (RJMCMC, Richardson and Green 1997) sampler has been developed in Sect. 4. The residual was modeled with a Gaussian process (GP) prior so as to reflect correlation across sites on a global scale.

Although not directly relevant to the modeling and data analysis in this article, we like to mention that there is a significant collection of literature on modeling extreme precipitation events. Here the quantity of interest is the right-hand tail of the rainfall distribution. The common approach is to use extreme value theory to propose a probability distribution for exceedance (the event that rainfall crosses a threshold value) rate at each site and relate the parameters of those distributions using spatial process models. See Cooley et al. (2007) and Sang and Gelfand (2009) for a hierarchical approach to this problem for rainfall data at point and grid-level, respectively.

Another important feature of our problem is prediction at thousands of unsampled sites. As Markov random field-based models do not have predictive property, one needs to include all such sites directly into the estimation, potentially leading to a large number of spatial random effects and slow convergence. With the spline-GP combination used in this article, prediction can be done as a post-MCMC analysis. One potential issue here is the sensitivity of GP computation to the number of locations in the dataset. There are a number of approximation techniques in the literature, such as process convolution (Higdon 2002), approximate likelihood (Stein et al. 2004), fixed rank kriging (Cressie and Johannesson 2008), covariance tapering (Furrer et al. 2006; Kaufman et al. 2008), predictive process (Banerjee et al. 2008) and very recently a combined approach involving reduced rank approximation and covariance tapering by Sang and Huang (2012); see Sun et al. (2012) for a review of available methods. We employ the fixed knot-based predictive process approximation, discussed in brief in Sect. 4.1. Finally, as with any Bayesian approach, information on uncertainty about process parameters, in addition to their pointwise estimates can be readily obtained which may be of significant interest with rainfall being a highly variable event.

The article is organized as follows. Measurements of precipitation around the world with multiple satellites and subsequent processing of the raw data are discussed in Sect. 2. In Sect. 3, the formulation of a Bayesian hierarchical spatial model for rainfall is introduced. Implementation details including the choice of priors, large data computation and sampling scheme are outlined in Sect. 4. The details of the simulation study and the real data analysis are presented in Sect. 5. Finally, Sect. 6 summarizes the present work and points out further related research. In this article, the notations $N(\mu, \sigma^2)$ and $LN(\mu, \sigma^2)$ have been used for denoting normal and lognormal probability densities with location μ and scale σ, respectively. Φ is used for the standard normal cumulative distribution function and $N_d(\mu, \Sigma)$ stands for d-dimensional multivariate normal distribution with mean vector μ and dispersion matrix Σ.

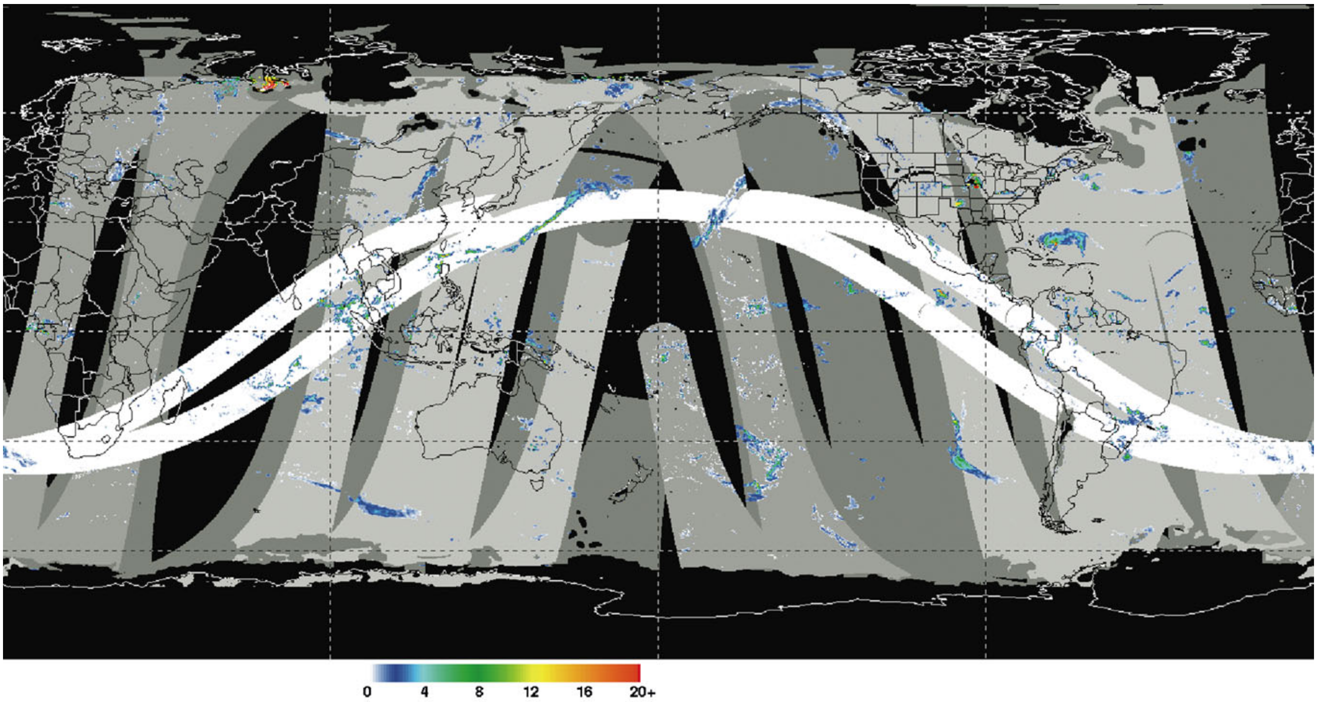


Fig. 1 Observation swaths of different satellites in a typical 3-hour observing period. The figure is from Huffman et al. (2007). *Black* indicates no data. *Colors* indicate rainrate in mm/hr. *Shades of gray* indicate observations of zero rain from the different satellites. The *white trajectory* corresponds to the path of TMI. Observations are averaged where overlaps occur. Reproduced by permission of the American Meteorological Society (Color figure online)

2 Collection of satellite measurements on precipitation

Precipitation observations from multiple satellites can be merged to yield global precipitation estimates at reasonably high spatial and temporal resolution as is done with the Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis (TMPA) described by Huffman et al. (2007). Data are collected by a variety of low earth orbit satellites, including the TRMM Microwave Imager (TMI) on the TRMM satellite, Special Sensor Microwave Imagers on the Defense Meteorological Satellite Program satellites, the Advanced Microwave Scanning Radiometer on the NASA *Aqua* satellite, and Advanced Microwave Sounding Units (AMSU) on National Oceanic and Atmospheric Administration (NOAA) operational satellites. These satellites vary with respect to altitude, orbital inclination, and equator crossing time. Data from multiple satellites are aggregated in 3-hour windows, which are approximately two complete orbits for low-Earth-orbiting satellites. The 3-hour time window is usual in precipitation study (Huffman et al. 2007) as a compromise between two competing goals: higher temporal resolution of the pattern of precipitation and greater spatial coverage during each averaging window. Because of the constraints of orbital motion and instrument design, a single satellite provides limited coverage of the globe within a 3-hour window. The current suite of operational and research satellites can, together, provide nearly global coverage in a 3-hour period. Longer sampling windows would offer greater coverage but would have a lesser ability to resolve the diurnal cycle, which is an important component of precipitation variation. Figure 1 shows the available data for a typical 3-hour period. For example, the regional 3-hour precipitation dataset we used for analysis in Sect. 5 has information from two satellites—TMI and AMSU-NOAA17. Following the TMPA practice, we choose the TMI as the reference standard due to its higher spatial resolution (resulting from relatively lower altitude) and ongoing calibration with the TRMM Precipitation Radar.

A typical satellite microwave radiometer measures the upwelling microwave radiation by using a rotating antenna to scan in a conical pattern as the satellite moves above the Earth's surface. The spatial resolution of the measurements depends on the altitude of the satellite, the size of the microwave antenna (typically about 1 m), and the wavelengths used. Rainrates are estimated using physical algorithms (Wilheit et al. 1977) based on reverse lookup tables that are precomputed using radiative transfer models. For efficiency and convenience in processing the data, within each 3-hour window the rain estimates from the instantaneous fields of view (pixels) of the satellite are first averaged onto a 0.25° × 0.25° latitude-longitude grid. In the tropics these boxes are nearly square and are approximately $28 \times 28\ \text{km}^2$.

This resolution represents a reasonable compromise among the differing spatial resolutions of the various instruments and is suitable for prediction related to climatological or agricultural studies. Rain measurements are reported as instantaneous rainrates in mm/hr.

3 Hierarchical spatial model for rainrate

In this section we describe the proposed hierarchical model for the multi-satellite precipitation data. Let D be the domain of observation, T be the time window of study, $X(s) = (x_1(s), x_2(s), \dots, x_p(s))^T$ be the p-dimensional vector of covariates measured at location s and $\tilde{Y}_l(s)$ is the observed rainrate during T from the l-th satellite at location $s \in D$ for $l = 1, 2, \dots, L$. However, at any given s, the rainrate may not be available for some or all of the L satellites. The observed rainrate data is modeled as a noisy version of an underlying unobserved potential rainrate process as follows:

$$\tilde{Y}_l(s) = \begin{cases} 0 \quad \text{with probability } \pi(s), \\ \exp\{c_{1l} + c_{2l}Y(s) + \varepsilon_l(s)\} \\ \quad \text{with probability } 1 - \pi(s), \end{cases} \tag{1}$$

where $\pi(s)$ is the probability of zero rainfall at grid cell s and $\exp(Y(s))$ represents the latent potential rainrate process at location s. If it rains, then the rainrate observed by satellite l, $\tilde{Y}_l(s)$, is a noisy measurement of that latent process. The parameters c_{1l}, c_{2l} are the model's additive and multiplicative bias adjustments specific to satellite l. For identifiability purpose, we need to set $c_{1l_0} = 0$, $c_{2l_0} = 1$ for some $l_0 \in \{1, 2, \dots, L\}$, so that any inference from the model can be interpreted with satellite l_0 as the reference. Generally, one chooses l_0 to be the satellite which is known to have maximum precision in measurements. In applications, where rainfall data are available from multiple sources, the one expected to have the highest accuracy can be used as the reference, e.g., raingauge data in Fuentes et al. (2008). The zero-mean noise $\varepsilon_l(s)$ characterizes the variations due to measurement error and/or micro-scale spatial variations for the l-th satellite.

We introduce the data augmentation approach (Tanner and Wong 1987; Albert and Chib 1993) to relate the rainfall probability $\pi(s)$ with the latent rainrate process $Y(s)$ in a flexible way. The spatial probit model for $\pi(s)$ is as follows:

$$\pi(s) = 1 - \Phi\big\{\mu_\pi(s) + \beta_\pi Y(s)\big\}. \tag{2}$$

Conceptually, this amounts to modeling the zeros of rainrate measurements to correspond to the low values of the latent process Y. The variable intercept $\mu_\pi(\cdot)$ can be referred to as "bias function" that accounts for the potential event of zero rainfall due to nonlinear interactions that cannot be captured linearly through $Y(s)$. If $\mu_\pi(s)$ is a constant over D, we have that $E\{\log \tilde{Y}_k(s) | \tilde{Y}_k(s) > 0\}$ and $\Phi^{-1}\{\pi(s)\}$ are linear functions of each other for any k, similar to the assumption made in Fuentes et al. (2008). We introduce L latent surfaces $Z_1(s), Z_2(s), \dots, Z_L(s)$ such that $Z_l(s) \overset{i.i.d.}{\sim} N(\mu_\pi(s) + \beta_\pi Y(s), 1)$, $1 \le l \le L$. Then we can rewrite (1) as

$$\tilde{Y}_l(s) = \begin{cases} 0 & \text{if } Z_l(s) \le 0, \\ \exp\{c_{1l} + c_{2l}Y(s) + \varepsilon_l(s)\} & \text{if } Z_l(s) > 0. \end{cases} \tag{3}$$

Estimation of $\{Y(s) : s \in D\}$ is of prime interest in this problem. $Y(s)$ captures rainfall patterns over D. Since rainfall over a small time window is a highly localized event, the usual isotropic GP-based spatial models will not suffice for Y. There are different approaches to introduce nonstationarity as outlined in Sect. 1. The approach we take here is to specify the mean surface $\mu_y(\cdot)$ using multivariate adaptive regression splines (MARS; Friedman 1991; Denison et al. 1998). The idea is to model the function as a sum of interactions of varying order from a basis set of local polynomials as follows:

$$Y(s) = \mu_y(s) + w(s),$$

$$\mu_y(s) = m_y(s) + \sum_{h=1}^{k_y} \nu_{y,h}\phi_{y,h}\big(X(s)\big), \tag{4}$$

$$\phi_{y,h}(x) = \prod_{r=1}^{n_h} \big[u_{hr}(x_{v_{hr}} - t_{hr})\big]_+,$$

where $m_y(\cdot)$ represents a fixed trend function (e.g. just an intercept or a linear or quadratic trend in components of s) and $(\cdot)_+ = \max(\cdot, 0)$, n_h is the degree of the interaction of the basis function $\phi_{y,h}$. The $\{u_{hr}\}$, sign indicators, are ±1, the v_{hr} gives the index of the predictor variable which is being split at the value t_{hr} within its range. Thus, the function $\phi_{y,h}$ represents a pattern around the knot $\boldsymbol{t}_h = \{t_{hr} : r = 1, 2, \dots, n_h\}$ (we refer to it as a *sp-knot*) and the set $\{\phi_{y,h}(\cdot) : h = 1, 2, \dots, k_y\}$ defines an adaptive partitioning of the multidimensional space.

Specifying $\mu_y(\cdot)$ with local interactions provides greater flexibility to model surface patterns and variable relationships. Most importantly, localized structures allow for nonstationarity. The flexibility of MARS lies in the fact that the interaction functions can be constructed adaptively, i.e., the order of interaction, knot locations, signs and even the number of such terms k_y is decided by the pattern of the data during model-fitting, eliminating the need for any prior adhoc or empirical judgement. In spite of having a flexible mean structure, MARS is relatively simple to fit which is an obvious advantage against other choices of nonstationary processes in the present application. For interpretability and to avoid overfitting, interactions up to a certain order are used and only up to a prefixed number of terms may be allowed in the above sum. Any additional or higher order global pattern is accounted for by assigning a zero-mean GP prior to the residual process $w(s)$. The covariance function for w is assumed to be isotropic, i.e. for two locations s and s', $\text{cov}\{w(s), w(s')\} = \sigma_y^2 \rho\{d(s, s'), \kappa\}$ where $d(\cdot,\cdot)$ is the great circle distance for latitude-longitude data. ρ is the correlation function (validity of usual correlation functions such as exponential and Matérn with respect to geodetic distance is discussed in Banerjee (2005)), κ is the parameter (vector) that controls the smoothness and rate of decay. The above specification implies that, for a set of n locations $\mathbf{s} = (s_1, s_2, \dots, s_n)$, the vector $Y(\mathbf{s}) = (Y(s_1), Y(s_2), \dots, Y(s_n))^T$ is distributed as:

$$Y(\mathbf{s}) \sim N_n\big(\mu_y(\mathbf{s}), \Sigma(\mathbf{s}; \kappa)\big),$$

where $\Sigma(i, i') = \sigma_y^2 \rho\{d(s_i, s_{i'}), \kappa\}$. If a priori $\boldsymbol{\nu}_y \sim N_{k_y} \times (\mathbf{0}, cI_{k_y})$ then marginalizing out $\boldsymbol{\nu}_y$, we have:

$$E\big[Y(\mathbf{s})\big] = m_y(\mathbf{s}), \text{V}\big[Y(\mathbf{s})\big] = cP(\mathbf{s})P(\mathbf{s})^T + \Sigma(\mathbf{s}, \kappa),$$

where $P(\mathbf{s})$ is a $n \times k_y$ matrix with i-th column being $(\phi_{y,i}(s_1), \phi_{y,i}(s_2), \dots, \phi_{y,i}(s_n))^T$. Let $\boldsymbol{\theta}_h = \{(v_{hr}, t_{hr}, u_{hr})$: $r = 1, 2, \dots, n_h\}$ be the parameters in $\phi_{y,h}$ and $f(s; \boldsymbol{\theta}_h) = \phi_{y,h}(s)$. Then at individual location level, we have:

$$\begin{aligned} \text{var}\big[Y(s_i)\big] &= c\sum_{h=1}^{k_y} f^2(s_i; \boldsymbol{\theta}_h) + \sigma_y^2, \\ \text{cov}\big[Y(s_i), Y(s_{i'})\big] &= c\sum_{h=1}^{k_y} f(s_i; \boldsymbol{\theta}_h) f(s_{i'}; \boldsymbol{\theta}_h) \\ &\quad + \sigma_y^2 \rho\big\{d(s_i, s_{i'}), \kappa\big\}. \end{aligned} \tag{5}$$

This provides a very flexible model for the covariance specification of the Y process that is also easy to interpret. The total covariance is decomposed into a locally varying (nonstationary) component combined with a global pattern coming from the GP. The vector $\boldsymbol{\theta}_h$ controls the characteristics of the pattern associated with the h-th sp-knot and $f(s, \boldsymbol{\theta}_h)$ represents its effect at location s. The resulting covariance is the sum of such local effects. In practice, one starts with a global covariance term *only* and then the model itself selects the local effects as necessary. The parameter c represents prior confidence (uncertainty) in these effects. The bias function, $\mu_\pi(\cdot)$ is also specified using MARS as above, with a separate set of parameters.

It follows from (5) that MARS specification amounts to building the covariance model of the output process with locally-supported components. In spatial literature, common approaches to incorporate nonstationarity by kernel mixing of process variables offers essentially similar decomposition of covariance; see Higdon et al. (1999), Fuentes (2002) and Banerjee et al. (2004). However, MARS offers significantly greater flexibility over those methods as the shape of such effects around the knots can be decided independently of each other without being controlled by that of any chosen kernel function. Moreover, allowing each of these patterns to have its own local support encourages sparsity and avoids the complexity often associated with determining appropriate kernel bandwidth parameter(s).

Another important advantage that this specification offers is the ability to let the required number of such local effects and the associated sp-knots be entirely decided by the data, without compromising for the computational complexity and interpretability. For any spatial model dependent on a set of knots, selection of an appropriate number of knots and placing them optimally over space has always been a critical issue. With too few knots, there is always a possibility of overestimating the actual spatial range and neglecting important local patterns. On the other hand, using too many knots increases the computational demand and may lead to poor predictive performance by accounting for even the noises or negligible variations in the observed data. Conditional on a fixed number of knots, Gelfand et al. (2012) discussed an approach to place them optimally using a minimum predictive variance criterion. A model-based approach for random knots was introduced in Guhaniyogi et al. (2011). There, in a multi-stage structure, a point process prior was assumed for the set of knots. The intensity of that point process can either be a parametric multimodal surface or a log-Gaussian process itself. However, when the number of such knots is significant, efficiently updating them may turn out to be challenging owing to a nonstandard posterior distribution. Our specification allows for random knot selection by placing a prior on the set of observed points in the space and then, during the MCMC scheme, varies the size as well as the members of the collection of the sp-knots via addition, deletion or modification, only one at a time. Another very important feature of our knot selection procedure is that even though we are working in a p-dimensional predictor space, a typical sp-knot can be a location in a lower dimensional space. This can be particularly useful when the spatial process under consideration has different degrees of smoothness across coordinates. For example, an atmospheric process may exhibit significant variations with change in longitude, but may remain relatively uniform with variations in latitude (at a fixed longitude). In those situations, a more parsimonious representation can be ensured with MARS by allowing the data to position its sp-knots only over the range of longitudes. Since most knot-based spatial methods have not addressed this dimension-wise variation so far, it can provide a potentially interesting direction for future research in this field. For our multi-stage spatial model, the sampling procedure is described in Sect. 4.

4 Details of estimation and inference

In this section, we focus on how to implement the model in Sect. 3 on a (potentially massive) precipitation dataset. Since GP computation is sensitive to the data size, we begin with a suitable approximation method to make the model capable of handling the computation. Subsequently we describe the full hierarchical model used to fit the dataset and outline the estimation procedure via MCMC. Finally, we mention some of the quantities of interest which can be estimated by post-processing the posterior samples.

4.1 Knot-based approximation for large dataset

When the number of locations inside D gets large (in thousands), updating the latent spatial process parameters inside an MCMC becomes complicated due to its high-dimensional covariance matrix. We use a reduced rank representation of the original process, the *predictive* process, as developed in Banerjee et al. (2008). Below, we present the idea in brief.

Consider realizations from a zero-mean, unit-scale GP $w(\cdot)$ at a set of n locations $\mathbf{s} = (s_1, s_2, \ldots, s_n) \in D$ where n is large. The method proceeds by first choosing $m \ll n$ locations $\mathbf{s^0} = (s_1^0, s_2^0, \ldots, s_m^0)$ in D, to be referred to as *pp-knots*, and then replaces $w(\mathbf{s})$ by an approximate process $\tilde{w}(\mathbf{s}) = E[w(\mathbf{s})|w(\mathbf{s^0})] = Lw(\mathbf{s^0})$ where the matrix L is calculated from the correlation function ρ of the original process w. L depends on the correlation parameter(s) of ρ. If $m \ll n$, we gain in terms of computation time using the Sherman-Woodbury-Morrison (S-W-M) type formulae (Banerjee et al. 2008). However the accuracy of the approximation goes up with increasing m, so there has to be a trade-off. We prefer to use this method, as it is derived directly from the parent process without any need for ad-hoc choice of basis functions, is easy to interpret and has closed form analytic expressions. This approximation is coherent with the MARS model introduced in Sect. 3, where the resulting covariance was shown to depend on a set of random sp-knots. We like to mention that, conditional on a fixed number of pp-knots m, we can allow their locations to vary by using a point process prior on them as in Guhaniyogi et al. (2011). However, the hierarchical model described in Sect. 3 already includes an adaptive spatial function based on sp-knots and GP is used only as a model for the residual process. So, in this article, we choose to work with a fixed set of pp-knots only.

We introduce bias correction, a modification discussed in Finley et al. (2009). Since $\text{var}\{w(\mathbf{s}_j)\} > \text{var}\{\tilde{w}(\mathbf{s}_j)\}$ for each j, the predictive process is expected to underestimate the spatial variance and increase the variance of the pure error. The correction introduces an heteroscedastic independent error ε^* so that $\tilde{w}(\mathbf{s}) = Lw(\mathbf{s^0}) + \varepsilon^*$ and $\text{var}\{\tilde{w}(\mathbf{s}_j)\} = \text{var}\{w(\mathbf{s}_j)\}$ for any $j = 1, 2, \ldots, n$. Introduction of a bias correction term also facilitates computation when the GP under consideration is applied to a latent-stage response (as ours) that needs to be updated every iteration. The advantage of this method is illustrated in Sect. 4.3.

4.2 MCMC from the complete hierarchical model

We discuss parameter estimation using a MCMC scheme from the model in Sect. 3. Let $\{s_{l1}, s_{l2}, \ldots, s_{ln_l}\}$ be the locations in D at which rainrate measurements are available from the satellite l for $l = 1, 2, \ldots, L$. Let $\tilde{y}_{lj}$ and x_{lj} denote the rainrate measurement and available covariate information, respectively, at location s_{lj} by satellite l for $j = 1, 2, \ldots, n_l$. Let $\mathbf{s}$ denotes the pooled set of n distinct locations $\bigcup_{l=1}^{L}\{s_{l1}, s_{l2}, \ldots, s_{ln_l}\}$ and $\mathbf{s^0}$ the set of m pp-knots as above. For the joint set of locations $(\mathbf{s}, \mathbf{s^0})$, partition the spatial correlation matrix as $C_{n+m}(\kappa) = \begin{pmatrix} C_n(\kappa) & C_{n,m}(\kappa) \\ C_{n,m}(\kappa) & C_m(\kappa) \end{pmatrix}$, where the entries of C_{n+m} are unit scale correlation terms under correlation function $\rho(\cdot, \kappa)$. From Sect. 4.1, $L(\kappa) = C_{n,m}(\kappa)C_m^{-1}(\kappa)$. Then we can write the full hierarchical model as follows:

$$
\begin{aligned}
&\tilde{y}_{lj} \sim 1(z_{lj} < 0)\delta_0 + 1(z_{lj} > 0)LN\big(c_{1l} + c_{2l}\,y(s_{lj}), \sigma_0^2\big),\\
&z_{lj} \sim N\big(\mu_\pi(s_{lj}) + \beta_\pi y(s_{lj}), 1\big),\\
&y(\mathbf{s}) = \mu_y(\mathbf{s}) + \sigma_y \tilde{w}(\mathbf{s}),\\
&\tilde{w}(\mathbf{s}) = L(\kappa)w\big(\mathbf{s^0}\big) + \varepsilon^*(\mathbf{s}), \qquad (6)\\
&w\big(\mathbf{s^0}\big) \sim GP\big(\mathbf{0}, \rho(\cdot, \kappa)\big),\\
&\varepsilon^*(\mathbf{s}) \sim N_n\big(\mathbf{0}, \text{Diag}\big\{I_n - C_{n,m}(\kappa)C_m^{-1}(\kappa)C_{m,n}(\kappa)\big\}\big),\\
&\mu_d(\mathbf{s}) = m_d(s) + \sum_{h=1}^{k_d} \nu_{d,h}\phi_{d,h}\big(x(\mathbf{s})\big), \quad d = \pi, y.
\end{aligned}
$$

Regarding prior specification, we set $c_{1l_0} = 0$, $c_{2l_0} = 1$ for some l_0 as discussed in Sect. 3. For $l \neq l_0$ we use Gaussian priors centered at 0 and 1, respectively. For the regression coefficient β_π and variance parameter σ_0^2, we use Normal and inverse-gamma priors, respectively, for conjugacy of posterior distributions. We choose $\rho(\cdot, \kappa)$ to be the exponential correlation function with decay parameter κ. It can be easily shown that $\kappa \approx 3.0/R$ (R is the spatial range, i.e., the distance at which the correlation falls below 0.05) so κ can be specified using a prior idea about possible values of R. An Inverse Gamma (a_σ, b_σ) prior was used for the spatial variance σ_y^2. We chose the trend function $m_d(\cdot)$ to be a constant intercept for $d = \pi, y$ and, without loss of generality, merge it with the MARS predictors as a constant basis function $\phi_{d,0}(\cdot) \equiv 1$. Conditional on k_π and k_y, the number of basis functions in the expansion, we assign Gaussian priors to the coefficient vectors so that $\boldsymbol{\nu}_y \sim N_{k_y}(\mathbf{0}, \sigma_y^2\tau_y^2 I_{k_y})$ and $\boldsymbol{\nu}_\pi \sim N_{k_\pi}(\mathbf{0}, \tau_\pi^2 I_{k_\pi})$. We assign Inverse gamma priors to both scale parameters τ_y^2 and τ_π^2 to maintain conjugacy, so that we can draw them from respective Inverse gamma posterior conditional distributions.

For $d = \pi, y$, we can control the parsimony of the nonstationary function μ_d in three different ways: (1) changing the prior mean of k_d, (2) putting an order constraint on each $\phi_{d,h}$ and (3) setting a fixed threshold k_0 for maximum value of k_d. Accounting for the column of ones corresponding to the constant basis function, $(k_d - 1)$ is chosen to have a Poisson(λ_d) prior truncated to the right at k_0 to control the number of terms in the sum. For each $\phi_{d,h}$ we can either use a strict upper bound (e.g. allowing only functions upto second order) or choose a prior that puts small probability on a higher order basis function. The value of k_0 should be chosen based on our idea of the variability in y-surface. $k_0 = 0$ corresponds to the most parsimonious model—a perfectly stationary surface. Increasing k_0 will allow us to capture more and more local patterns but risks overfitting. In practical examples, the choice of k_0 may come from prior idea about nature of variation in the surface. In absence of any such information, one can use a validation method, i.e., using a subset of the data as the test set, fit the model with different values of k_0 and investigate its influence on the predictive performance.

During the MCMC, the vector of parameters has been updated in the following blocks: (i) spline coefficients $\{\nu_{d,h}\}$, number of basis functions k_d and parameters within each $\phi_{d,h}$ for $d = \pi, y$; (ii) latent rainrate variables $y(\mathbf{s})$; (iii) auxiliary surface $z(\cdot)$; (iv) regression parameters such as β_π and $\{c_{il} : i = 1, 2, l = 1, 2, \ldots, L, l \neq l_0\}$. We rewrite the probability distribution of $y(\mathbf{s})$ as $y(\mathbf{s}) \sim N_n(\mu_y(\mathbf{s}) + L(\kappa)w(\mathbf{s^0}), \Sigma_y)$ where $\Sigma_y = \sigma_y^2 \mathrm{Diag}\{I_n - C_{n,m}(\kappa) \times C_m^{-1}(\kappa)C_{m,n}(\kappa)\}$. The resulting posterior distribution for y is Gaussian and, *importantly*, has independence across locations. So, it is easy to draw y even when n is large. As in Albert and Chib (1993), conditional on $y(s_{lj})$ and $\tilde{y}_{lj}$, z_{lj} can also be sampled independently across different satellites and locations:

$$z_{lj}|\tilde{y}_{lj}, y(s_{lj}), \mu_\pi, \beta_\pi \overset{\text{ind}}{\sim} 1(\tilde{y}_{lj} = 0)N^{(-\infty,0)}(\mu_{lj}, 1) + 1(\tilde{y}_{lj} > 0)N^{(0,\infty)}(\mu_{lj}, 1)$$

$$\mu_{lj} = \mu_\pi(s_{lj}) + \beta_\pi y(s_{lj}),$$

where $N^A(\mu, 1)$ stands for $N(\mu, 1)$ distribution truncated within $A \subset \mathbb{R}$.

Next, we discuss updating parameters related to the posterior distribution of y, i.e., κ, σ_y^2 and μ_y. For this, we first marginalize out $w(\mathbf{s^0})$ from the distribution of y. As observed in Chib and Carlin (1999), marginalizing out the random effects improves the mixing behavior of the MCMC. However, this leads to a full covariance structure for y as:

$$\Sigma\big(y(\mathbf{s})\big) = \sigma_y^2\big[C_{n,m}(\kappa)C_m^{-1}(\kappa)C_{m,n}(\kappa) + \mathrm{Diag}\big\{I_n - C_{n,m}(\kappa)C_m^{-1}(\kappa)C_{m,n}(\kappa)\big\}\big].$$

We can use S-W-M type matrix computations for calculating the determinant and inverse of this covariance matrix and use that in drawing from posterior distributions of κ and σ_y^2. Posterior samples of $w(\mathbf{s^0})$ can be drawn afterwards from a multivariate normal distribution as evident from (6). Regression parameters c_{1k}, c_{2k} and β_π can also be updated using standard sampling steps. The most important component of this MCMC is updating the spline parameters appearing in μ_y and μ_π which has been performed using a reversible jump (Richardson and Green 1997) scheme described below.

We start with μ_y, the mean function for y. We drop the suffix y for notational simplicity. With k basis functions, let $\alpha_k = \{n_h, \mathbf{u}_h, \mathbf{v}_h, \nu_h, \mathbf{t}_h\}_{h=1}^k$ be the corresponding set of spline parameters. Marginalizing out ν and σ^2, the distribution for $y(\mathbf{s})$, $p(y(\mathbf{s})|k, \alpha_k, \ldots)$, can be written in closed form (see Appendix). Now, using a suitable proposal distribution q, propose a dimension changing move $(k, \alpha_k) \rightarrow (k', \alpha_{k'})$. We consider three types of possible moves (i) *birth*: addition of a basis function; (ii) *death*: deletion of an existing basis function; and (iii) *change*: modification of an existing basis function. Thus $k' \in \{k - 1, k, k + 1\}$. The acceptance ratio for such a move is given by

$$p_{k \rightarrow k'} = \min\left(1, \frac{p(y(\mathbf{s})|k', \alpha_{k'}, \ldots)}{p(y(\mathbf{s})|k, \alpha_k, \ldots)} \frac{p(\alpha_{k'}|k')p(k')}{p(\alpha_k|k)p(k)} \times \frac{q\{(k', \alpha_{k'}) \rightarrow (k, \alpha_k)\}}{q\{(k, \alpha_k) \rightarrow (k', \alpha_{k'})\}}\right).$$

First, we mention the prior for (k, α_k) in the form of $p(\alpha_k|k)p(k)$. As specified above, $(k + 1)$ has a Poisson(λ) prior truncated at some upper bound k_0. Within each constituent local polynomial, n_h controls its order, $\mathbf{v}_h$ controls the set of variables involved whereas $\mathbf{u}_h$ and $\mathbf{t}_h$ determine the signs and the position of the sp-knot. If p is the total number of covariates and we allow interactions up to 2nd order, then the number of possible choices for a basis function (excluding the constant function) is $N = p + p + \binom{p}{2} = \frac{p^2+3p}{2}$. Accounting for rearrangement of the same set of basis functions, the number of distinct k-set basis functions is $N^k/k!$. Once we determine a basis function, we choose the individual coordinates of that sp-knot uniformly from the available data points (since a change in pattern can only be detected at data points) and determine its sign to be positive or negative with probability 1/2 each. Thus we obtain:

$$p(\alpha_k|k) \propto \frac{N^k}{k!}(1/2n)^{\sum_h^k n_h}.$$

Although above we assumed that all covariates have n distinct values to locate a sp-knot, modifications can be made easily when this is not the case.

Next we specify the proposal distribution $q(\cdot,\cdot)$ for each of the three moves as follows:

(i) First decide on the type of move to be proposed with probabilities b_k (birth), d_k (death) and c_k (change), $b_k + d_k + c_k = 1$. We put $d_k = 0, c_k = 0$ if $k = 1, b_k = 0$ if $k = k_0$.

(ii) For a *birth* move, choose a new basis function randomly from the N-set, calculate its order n_h and choose the location of its sp-knot and signs as before with probability $(\frac{1}{2n})^{n_h}$.

(iii) The *death* move is performed by randomly removing one of the $k-1$ existing basis functions (excluding the constant basis function).

(iv) A *change* move consists of choosing an existing non-constant basis function randomly and alter its sign and corresponding sp-knot.

From above, we have

$$q\big((k,\alpha_k) \to (k',\alpha_{k'})\big) = \begin{cases} b_k \frac{1}{N}(\frac{1}{2n})^{n_{k+1}} & k' = k+1, \\ d_k \frac{1}{k-1} & k' = k-1, \\ c_k \frac{1}{k-1}(\frac{1}{2n})^{n_h} & k' = k. \end{cases}$$

Above, for the 'change' step, h denotes the index of basis function that has been randomly chosen for change. The acceptance ratios for different types of move can be worked out from this. Set $k = k', \alpha_k = \alpha_{k'}$ if the move is accepted, leave them unchanged otherwise. Subsequently, ν_y can be updated using the k-variate t distribution with degrees of freedom $d = n + 2a_\sigma$, mean μ_k, dispersion $\frac{c_{0k}\Sigma_k}{d}$, whose expressions are given (with derivation) in Appendix. The updating scheme for the spline parameters in μ_π is similar except for the fact that we need to marginalize over ν_π only as z has a known variance (set to 1) as in (6).

4.3 Posterior inference

The principal objective of this data analysis is to understand the precipitation pattern over the region D. For that, we create spatial maps of (i) expected rainrate $\{\pi(s)\exp\{Y(s)\} : s \in D\}$; and (ii) probability of rainfall $\{\pi(s) : s \in D\}$ using the posterior samples. There can be locations inside D with no available measurements from any of the L satellites that do not have any likelihood contribution. The realizations of the rainrate process Y at those locations are constructed using the predictive distributions of a GP. Since Gaussian processes can capture a wide range of dependencies, using them in a hierarchical setting enhances predictive performance for the model. Mathematically, if only n out of N sites have at least one satellite measurement then inference for the remaining $(N-n)$ sites is done from their posterior predictive distributions. If $\mathbf{s}_p = \{s_{n+1}, s_{n+2}, \ldots, s_N\}$ denotes the set of locations with no precipitation data, the foregoing predictive process approximation yields

$$\begin{aligned} &y(\mathbf{s}_p) = \mu_y(\mathbf{s}_p) + \sigma_y\big\{C_{N-n,m}(\kappa)C_m^{-1}(\kappa)w\big(\mathbf{s^0}\big) + \varepsilon^*(\mathbf{s}_p)\big\}, \\ &\varepsilon^*(\mathbf{s}_p) \sim N_{N-n}\big(0, \mathrm{Diag}\big\{I_{N-n} - C_{N-n,m}(\kappa)C_m^{-1}(\kappa) \\ &\qquad \times C_{m,N-n}(\kappa)\big\}\big). \end{aligned} \tag{7}$$

As we mentioned earlier, the advantage of using bias correction is evident here since conditional on $w(\mathbf{s^0})$ and κ, we can draw samples from the posterior predictive distribution of $Y(\mathbf{s}_p)$, independent of each other and also of $Y(\mathbf{s})$ (conditional on realizations of $w(\mathbf{s^0})$) due to the independence among ε^*s across locations. This is computationally very efficient since we do not need to draw from a high dimensional multivariate Gaussian distribution if we want to study a larger region and require predicting the rainrate surface at thousands of sites with no satellite readings. Also of interest is the posterior estimate of $\{\mu_y(s) : s \in D\}$, which provides an idea of localized patterns in the rainrate over D. All these diagnostics are provided with the real data analysis in Sect. 5.

5 Data analysis

We proceed to application of the variable-knot approach described in Sects. 3 and 4. First, in Sect. 5.1 we carry out simulation studies to highlight the improvement in predictive performance under the proposed method relative to fixed-knot predictive process models. Then, in Sect. 5.2, we analyze an actual precipitation dataset from Northern South America.

5.1 Simulation study

In the discussion following (4) and (5), we have argued that one of the key advantage of a MARS-based covariance model is its ability to capture a wide range of spatial structures in the response surface. To justify that numerically, we use synthetic datasets from two different models.

For simulation, we fix the input space $\mathbb{X}$ to be the unit hypercube in $\mathbb{R}^4$. The input points $\mathbf{x} = (x_1, x_2, x_3, x_4)^T$ are drawn uniformly over $\mathbb{X}$ and the response, $Y(\mathbf{x})$, is simulated from a Gaussian process with a nonstationary covariance function as follows:

$$\begin{aligned} &E\big[Y(\mathbf{x})\big] = \beta_0; \\ &\mathrm{Cov}\big[Y(\mathbf{x}), Y\big(\mathbf{x}'\big)\big] = \sigma_0^2 \mathbb{I}_{\mathbf{x}=\mathbf{x}'} + \mathbf{x}^T \Omega_0 \mathbf{x}', \end{aligned} \tag{A}$$

Table 1 Comparison of predictive performance for different spatial models

Simulation model	Prediction property	Model for estimation			
		MARS w/pp-knots	Predictive process with		
			same # pp-knots	2× # pp-knots	3× # pp-knots
(A)	Abs. Bias	0.099	0.104	0.107	0.112
	Pred. Uncertainty	0.570	0.626	0.603	0.603
	Coverage Propn.	92.100	91.500	91.000	90.600
(B)	Abs. Bias	0.048	0.271	0.195	0.137
	Pred. Uncertainty	0.539	1.937	1.536	1.374
	Coverage Propn.	96.200	93.500	93.800	94.500

where $\mathbb{I}_A = 1$ if A is true, 0 otherwise. For any $\sigma_0^2 > 0$ and any positive definite matrix Ω_0, it is easy to verify that the above is a valid covariance function.

In the second example, keeping $\mathbb{X}$ unchanged, we move to a more general functional form for $Y(\mathbf{x})$ as follows:

$$Y(\mathbf{x}) = \beta_0 + \beta_1 x_1^5 + \log\big(1 + x_2^2\big) + \beta_2 x_3 \sin\big(\pi x_1^2\big) + \mathbb{I}_{x_4 < 0.5}(x_4 - 0.2)^2 \exp(x_2 + 3) + (x_3 - 0.75)^2 x_4. \tag{B}$$

For comparison based on the above simulations we choose four competing models for $Y(\mathbf{x})$—(i) MARS with pp-knot based residuals as used in (6) and fixed-knot predictive process with (ii) equal, (iii) twice and (iv) thrice as many pp-knots as in (i). We note here that the models in (ii)–(iv) can be implemented as a special case of (i) by setting $\nu_h = 0$ for all h. From each of (A) and (B), we generate 1000 observations. Then we randomly drop 10 % of the points to create a 'test set', fit the four candidate models (with 30, 30, 60 and 90 pp-knots, respectively) on the remaining points. For each observation in the test set, we compute three measures of predictive performance: (i) absolute bias: the magnitude of how far the observed value of any Y is from its estimated value, (ii) predictive uncertainty: the width of the 90 % credible set for MC samples of any Y, i.e., the range of samples from the posterior distribution of Y excluding the smallest and largest 5 % of the draws and (iii) coverage status: whether a specific Y in our test set actually falls inside its 90 % posterior credible set. We replicate this procedure ten times randomizing selection of the test observations. Below, in Table 1, we summarize the results over all replications.

Table 1 shows that MARS with predictive process prior on residuals has performed considerably better than the other three for both simulation models in terms of all the criteria—bias, uncertainty and coverage rates. Most importantly, even with two to three fold increase in knot-size in the predictive process approximation, the prediction performance achieved with MARS cannot be attained. First, in model (A), where we are trying to fit observations coming from a nonstationary GP, MARS considerably brings down the prediction uncertainty for test samples without compromising on the coverage rates. Using predictive process only has produced wider prediction intervals to achieve comparable coverage rates, irrespective of whether knot size is doubled or tripled. More drastic difference in performance is observed in case of model (B), where $Y(\mathbf{x})$ has a general functional form. Use of sp-knots has significantly brought down the bias and uncertainty of prediction but it still achieves the highest coverage proportion among all three methods. Doubling the number of pp-knots (and subsequent tripling, too) helps in improving the quality of predictive process approximation but still lags far behind the one achieved with MARS-based model. Increasing the number of knots actually gives rise to the computational demand. Hence, with larger datasets, we expect to see even more significant gains in predictive accuracy of our variable knot approach compared to conventional GP approximations without sacrificing the computational efficiency.

At this point, looking at the flexibility of MARS-based variable-knot spatial structure, it is reasonable to ask whether we can replace the GP prior on the residual process $w(\cdot)$ in (4) with a white noise (WN) without significant loss of performance. This would simplify the MARS + GP specification for Y to a MARS + WN type model, thus eliminating the need to use predictive process approximation of Sect. 4 even when the number of observation is large. Actually, it can be shown that when MARS + GP is allowed to use fewer interactions (smaller k_0, hence a simpler nonstationary structure) than MARS + WN, even then it can lead to much improved prediction characteristics such as smaller bias and shorter predictive intervals than the latter. We refer the reader to Chakraborty et al. (2013, Sect. 5.1) for a discussion on this and other potential issues regarding application of MARS-based models.

5.2 Multi-satellite precipitation data

For a real-world application of the proposed method, we consider the rainrate data from a region in northern South

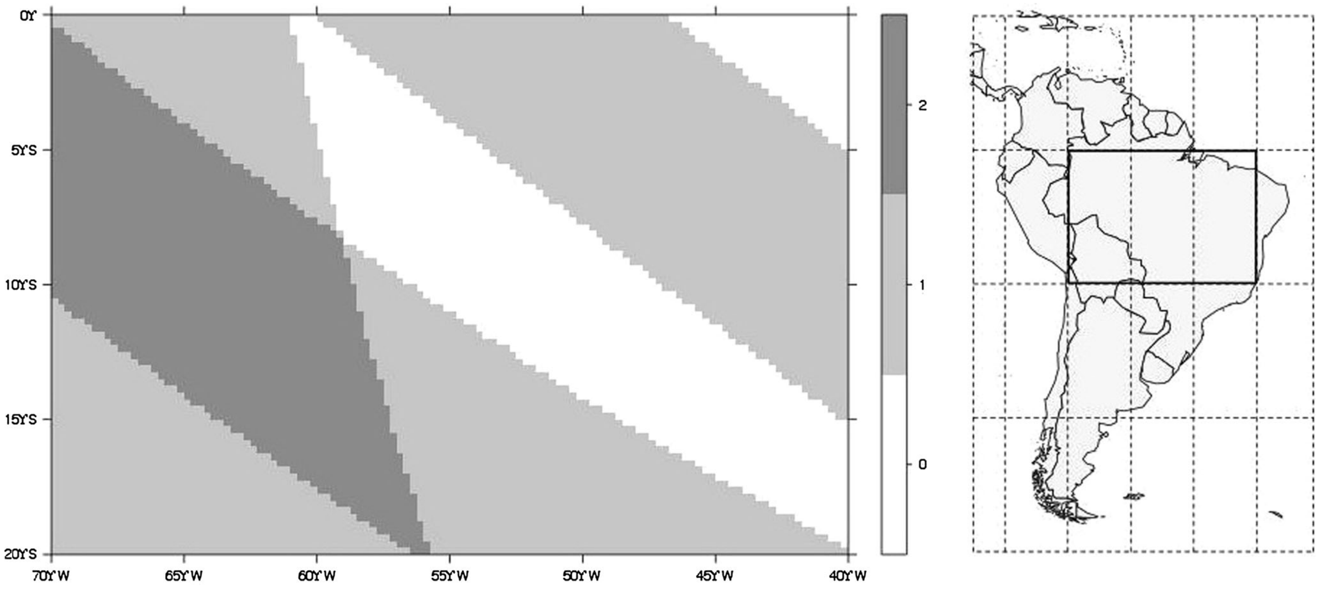


Fig. 2 (*Left*) Map of the study region showing number of satellite measurements available across different sites during 1.30 a.m.–4.30 a.m. on January 1st, 2008. (*Right*) Location of the study region in South America

America (Fig. 2) lying south of the equator, inside the rectangle $[70° W, 40° W] \times [20° S, 0°]$. This region is chosen because of its large average rainfall rate, which helps to reduce sampling errors, and its regular diurnal variability. A long term mean rainrate pattern over 13 years for this region is presented in Fig. 3. In the current work, we select a 3-hour time window, 1.30 a.m.–4.30 a.m. on January 1st, 2008. Data are available from two satellites—TMI and AMSU-NOAA17. Other satellites did not make any observation in this region during the specified time interval. Ground-based observations by gauges or radar are very sparse in the Amazon basin. From Fig. 2, it is evident that some of the sites have multiple rainrate measurements as they fell inside the intersection of trajectories of both satellites during that time window whereas some parts of the region were not covered by any of the satellites. Thus, estimation of rainrate at observed locations, from either one or both of the satellites as well as prediction at unmapped sites are necessary to create a complete precipitation map for this region.

We start with some empirical summaries of the raw data. Rainrate measurements from both of the satellites are available at 2026 of a total of 9600 sites in the region whereas 2305 of them have no available data. During the time window, TMI was able to cover 5653 sites whereas AMSU-NOAA17 covered 3668 sites. About 17 % of all satellite measurements recorded positive precipitation, the rest being all zero. Figure 4 provides a comprehensive representation of the precipitation measurements collected during that time interval from the above region.

We analyze the data using the model from (6). We use diffused prior specifications for model parameters. Regression coefficients such as β_π, c_{11} and c_{21}, are assigned a $N(0, 100)$ prior. For variance parameters, we use an Inverse Gamma(2, 4) prior. For the sp-knots in μ_y and μ_π, we allow each of them to have at most $k_0 = 30$ nonconstant local functions, i.e. each of k_y and k_π is assigned an $1 + \text{Poisson}(4)$ prior truncated between $[1, 31]$. As we show later, this range turns out to be adequate for our dataset. We consider local polynomials of only upto second order, i.e. n_h has a uniform prior on $\{1, 2\}$. We select $m = 100$ locations within the region as the pp-knots. For the correlation parameter κ, we first select a possible set of values for the spatial range R and use a uniform distribution over equidistant points in that set. As mentioned in Sect. 4.2, this leads to a discrete uniform prior for κ. The MCMC is run for 15000 iterations, discarding the first 5000 draws and thinning the rest at every 5th draw.

Before providing the posterior summaries for precipitation patterns, we perform a validation step by comparing model-based predictions with corresponding true observations. For that, we randomly remove a "test" set that consists of 130 and 122 sites from the region with positive precipitation records from TMI and AMSU-NOAA17, respectively. We treat those sites as having no available measurement, predict the latent process Y as in Sect. 4.3 and, tracing back the hierarchy in (6), regenerate the samples for the "observed" rainrates $\tilde{Y}$ at each of those locations for each of the satellites using 2000 thinned MC samples of model parameters. The samples are summarized in form of predictive mean and credible set. As with the simulation studies, we compute three measures of predictive performance: (i) absolute bias (ii) predictive uncertainty and (iii) coverage status for every $\tilde{Y}$ in the test set. In Table 2 we present, for each

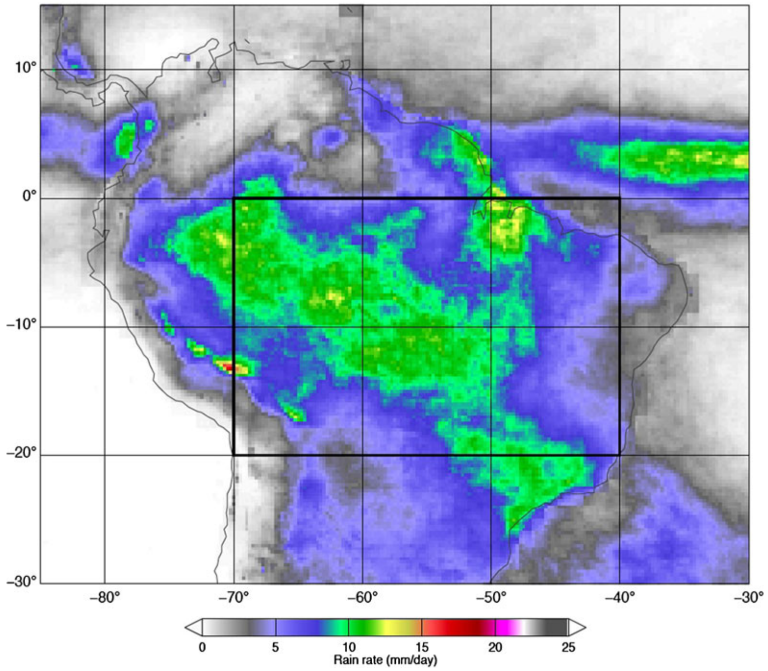


Fig. 3 Climatological mean surface rainrate for January for the period 1998–2010 from the TRMM Multisatellite Precipitation Analysis (Huffman et al. 2007). The domain for this study is indicated by the *black rectangle box*. The mean rainrate field is relatively smooth across the study domain, except in the southwest corner, where orographic effects from the Andes Mountains are apparent (Color figure online)

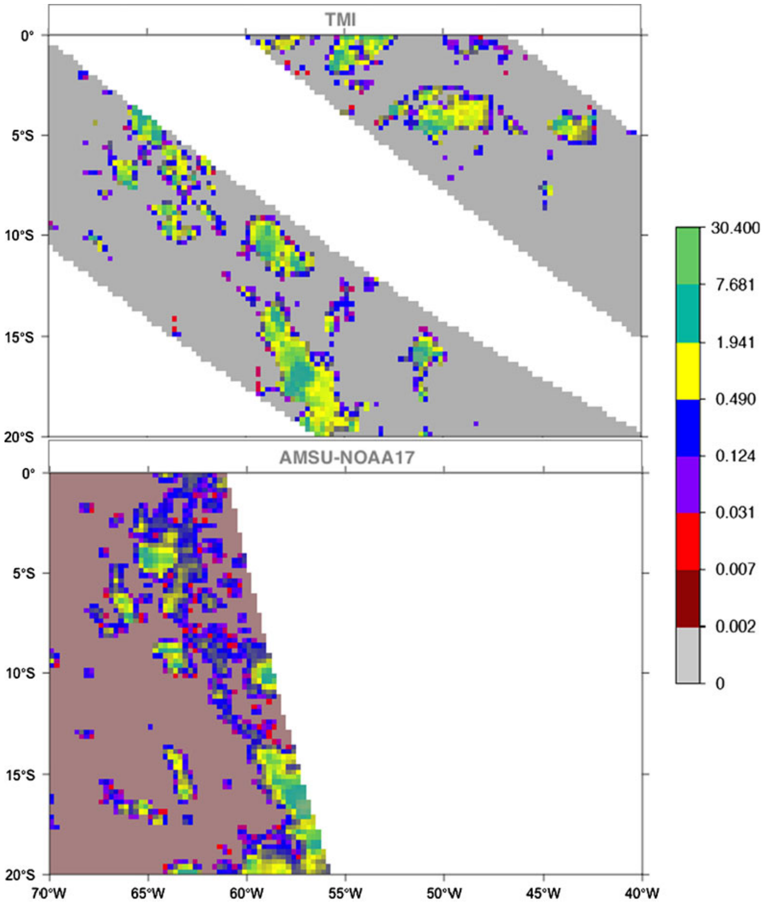


Fig. 4 Grid-level rainrate measurements from (*top*) TMI and (*bottom*) AMSU-NOAA17 satellites during [1:30, 4:30] a.m. on January 1, 2008 (Color figure online)

satellite, the summary of these measures—the mean absolute bias, the mean predictive uncertainty and the coverage proportion.

The results turn out to be satisfactory as the empirical coverage rates for both satellites exceed 89 %. As expected, the TMI measurements have significantly better estimates of error and reduced uncertainty of prediction over those of AMSU-NOAA17 because the former has been used as the reference standard for this data analysis. Now, we provide posterior summary statistics for some important model parameters in Table 3.

Table 2 Predictive performance of both satellites on test dataset

Satellite	Mean absolute bias (in mm/hr)	Mean predictive uncertainty (in mm/hr)	Coverage proportion for 90 % credible sets
TMI	1.039	3.765	0.892
AMSU-NOAA17	1.697	5.738	0.893
Combined	1.357	4.720	0.893

The number of sp-knots for the latent log-rainrate process as well as for the probability of no precipitation are well-within their assigned range of [1, 31]. The multiplicative factor for the AMSU-NOAA17, c_2 is marginally above 1. The effect of the latent log rainrate process Y on the event of rainfall is parametrized by β_π and turns out to be significant. Next, we look at the spatial maps for rainrate summaries. First, we present the pointwise estimates for the exp[Y] and π surfaces in Fig. 5. Then, the posterior mean and uncertainty estimates (90 % credible set width) of the expected rainrate process, as defined in Sect. 4.3, are included in

Table 3 Posterior summaries of important model parameters

Parameter type	Parameter name	Point estimate	90 % posterior credible interval
Latent process-specific	k_y	21	[17, 26]
	k_π	8	[5, 12]
	β_π	1.392	[1.271, 1.582]
AMSU-NOAA17-specific	c_1	0.332	[0.245, 0.429]
	c_2	1.090	[0.999, 1.187]

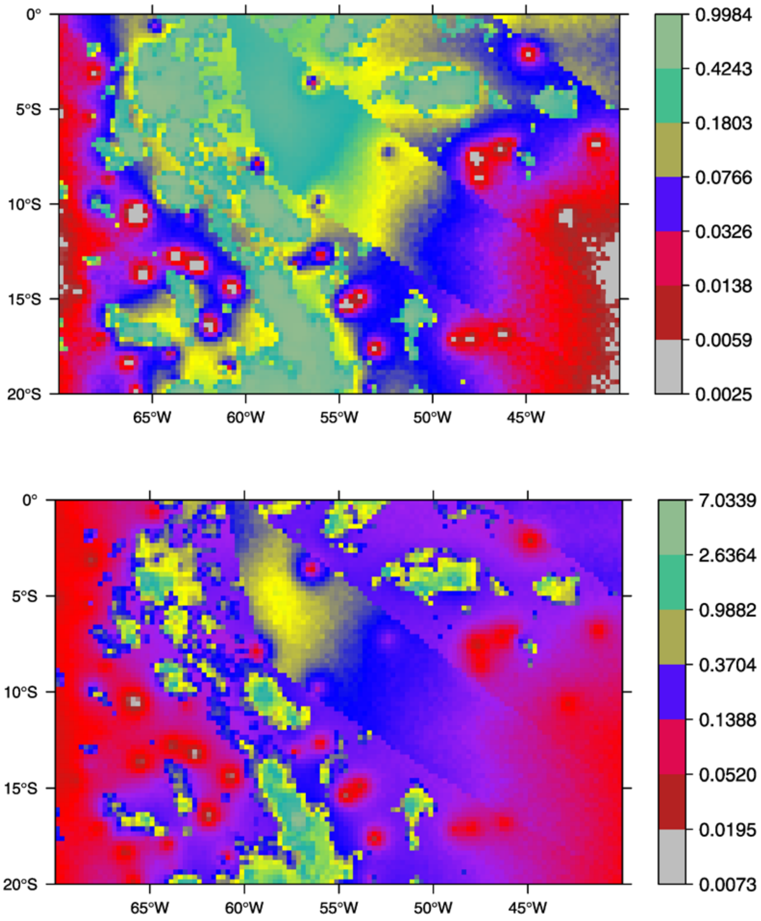


Fig. 5 Posterior surface estimates of (*top*) probability of rainfall π and (*bottom*) potential rainrate exp[Y] (Color figure online)

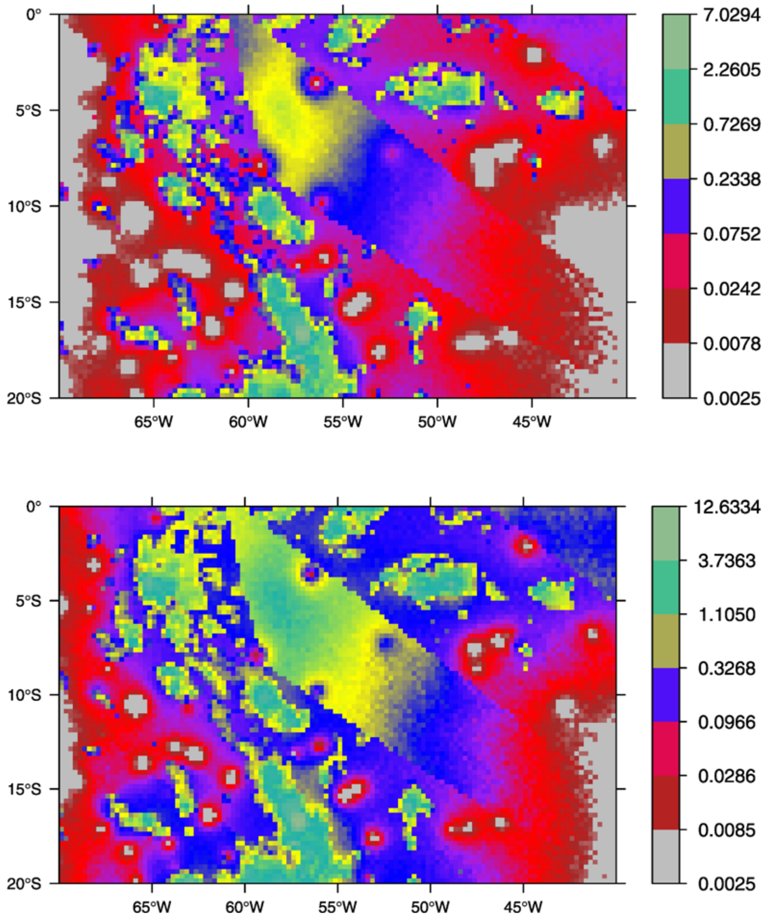


Fig. 6 Posterior estimated surfaces for (*top*) expected rainrate and (*bottom*) its uncertainty (width of pointwise 90 % credible sets) (Color figure online)

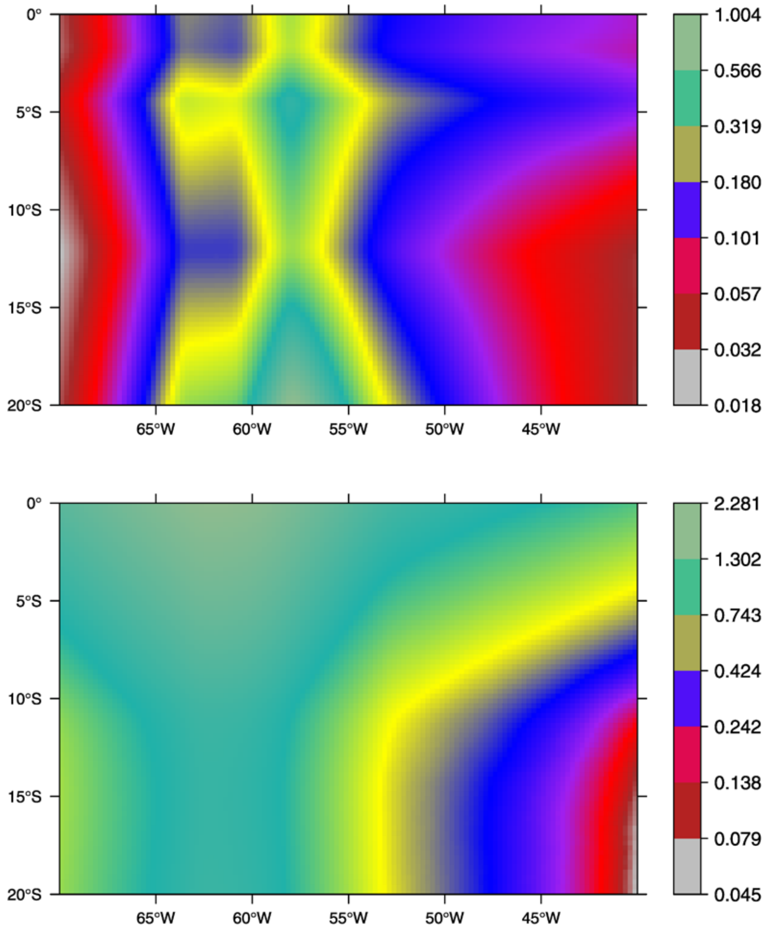
Fig. 6. To highlight the contribution of sp-knot based functions, we present the posterior maps of $\exp\{\mu_y(\cdot)\}$ and μ_π in Fig. 7.

Precipitation is mostly concentrated in the west-central part of the region and decreases as one moves towards the ocean in the east. In Fig. 5, the probability map shows higher chance of observing precipitation in the central part of the region as a result of smoothing effect of high rainfall observations in the surrounding regions from both satellites. Figure 6 shows patches of region with relatively higher expected rainrate. The highs and lows of the uncertainty estimates are often related to those of the corresponding point estimates. This is a natural property of the log-Gaussian model (also other models like Gamma) due to the interdependence between mean and variance. The uncertainty estimates are reflective of typically high variability (as well as lack of spatial smoothness) of precipitation over a short time window. Finally, Fig. 7 shows bands of piecewise homogeneous regions created by the collection of sp-knots and associated local polynomials. The estimated surface for $\exp[\mu_y]$ contains relatively higher number of localized patterns than the surface for μ_π. This is justified by Table 3, which shows that the posterior probability mass function for k_y puts greater weight on larger values within the $[1, 31]$ range than the one for k_π.

6 Summary and future work

In this article, we presented a novel hierarchical model to combine precipitation measurements from multiple satellites. To capture a wide range of localized as well as large-scale spatial patterns in the underlying potential rainrate process, a flexible random knot-based mean function has been used in combination with a stationary residual in the log scale. The method was adjusted to handle a large number of observation locations using a predictive process approximation, making it applicable to studies involving larger regions. However, it is likely for a larger domain to contain heterogeneous subregions, e.g., land-sea boundaries or regions separated by mountain stretches, that may experience different rainfall patterns. Whereas the MARS specification, employed in this work, can be really useful for these situations (since it does not make assumptions regarding any global pattern of correlation in the response surface), it is worth exploring alternative modeling ideas that can account

Fig. 7 Maps of sp-knot based surfaces—(*top*) exp[μ_y] and (*bottom*) μ_π constructed from their respective posterior samples (Color figure online)



for boundary effects between nonhomogeneous regions. Another extension lies in extending the 3-hr window to a larger time interval like a day or a week that brings a temporal pattern to the data. The associated spatio-temporal process can be developed as an extension to the current spatial model for Y in a number of different ways. Limitations for such specifications with respect to restricted model assumptions (e.g. separability across space and time), computational load or intuitive interpretability need to be compared for preferring one of them over the others. Another important information that may significantly improve rainfall prediction is the use of associated covariate data. Covariates can be of different types: (i) climate features such as temperature, wind speed, wind direction etc., (ii) geographic information such as elevation and (iii) measures of human intervention, e.g., forest cover, emission rates of pollutants etc. which are believed to influence rainfall in the long run. Inclusion of appropriate local covariate information is often useful to explain the nonstationarity, thus eliminating the need for complex models.

Acknowledgements The authors acknowledge the Texas A&M University Brazos HPC cluster that contributed to the research reported here (http://brazos.tamu.edu).

Appendix: Marginalizing out ν_y and σ_y^2 for estimation of spline parameters in $\mu_y(s)$

Denote by $\ldots$ all parameters except ν, σ_y^2. Let $P = [\phi_1[x(\mathbf{s})], \phi_2[x(\mathbf{s})], \ldots, \phi_k[x(\mathbf{s})]]$, $S = y(\mathbf{s}) - P\nu_y$. We have,

$$
\begin{aligned}
&p\big(y(\mathbf{s})|\ldots\big)\\
&\quad\propto \int_{\nu_y}\int_{\sigma_y^2} p\big(y(\mathbf{s})|\nu_y, \sigma_y^2, \ldots\big)\, p\big(\nu_y|\sigma_y^2\big)\, p\big(\sigma_y^2\big)\, d\sigma_y^2\, d\nu_y,\\
&\quad\propto \big(2\pi\tau_y^2\big)^{-k/2}\int_{\nu_y}\int_{\sigma_y^2}\big(\sigma_y^2\big)^{-\frac{n+k}{2}-a_\sigma-1}\\
&\qquad\times \exp\left[-\frac{1}{2\sigma_y^2}\big(S^T D^{-1} S + \nu_y^T\nu_y/\tau_y^2 + 2b_\sigma\big)\right] d\sigma_y^2\, d\nu_y,\\
&\quad\propto \big(2\pi\tau_y^2\big)^{-k/2}\,\Gamma\left(\frac{n}{2}+a_\sigma\right)\\
&\qquad\int_{\nu_y}\left(\frac{S^T D^{-1} S + \nu_y^T\nu_y/\tau_y^2}{2} + b_\sigma\right)^{-\frac{n+m+k}{2}-a_\sigma} d\nu_y.
\end{aligned}
$$

Now write $S^T D^{-1} S + \nu_y^T \nu_y = \nu_y^T A \nu_y - 2\nu_y^T B + C$, where $A = P^T D^{-1} P + \frac{I_k}{\tau_y^2}$, $B = P^T D^{-1} S_y$, $C = S_y^T D^{-1} S_y$. Then we have, $S^T D^{-1} S + \nu_y^T \nu_y + 2b_\sigma = (\nu_y - \mu_k)^T \Sigma_k^{-1} (\nu_y - \mu_k) + c_{0k}$, where $\mu_k = A^{-1} B, \Sigma_k = A^{-1}, c_{0k} = C - b^T A^{-1} b + 2b_\sigma$. Denote $d = n + 2a_\sigma$. Then

$$
\begin{aligned}
&p\big(y(\mathbf{s})|\ldots\big) \\
&\quad \propto \left(\pi \tau_y^2\right)^{-k/2} c_{0k}^{-\frac{d+k}{2}} \Gamma\left(\frac{d+k}{2}\right) \int_{\nu_y} \left[\frac{1}{d}(\nu_y - \mu_k)^T\right. \\
&\qquad \left. \times \left(\frac{c_{0k} \Sigma_k}{d}\right)^{-1} (\nu_y - \mu_k) + 1\right]^{-\frac{d+k}{2}} d\nu_y .
\end{aligned}
$$

The integrand is the pdf (up to a constant) for the k-variate t distribution with mean μ_k, dispersion $\frac{c_{0k} \Sigma_k}{d}$ and degrees of freedom d. Hence, we obtain the closed form expression for $p(y(\mathbf{s})|\ldots) \propto (\tau_y^2)^{-k/2} c_{0k}^{-\frac{d}{2}} |\Sigma_k|^{1/2}$.

References

Agarwal, D.K., Gelfand, A.E., Citron-Pousty, S.: Zero-inflated models with application to spatial count data. Environ. Ecol. Stat. **9**, 341–355 (2002)

Albert, J.H., Chib, S.: Bayesian analysis of binary and polychotomous response data. J. Am. Stat. Assoc. **88**(422), 669–679 (1993)

Anderes, E.B., Stein, M.L.: Estimating deformations of isotropic Gaussian random fields on the plane. Ann. Stat. **36**, 719–741 (2008)

Austin, P.M., Houze, R.A.: Analysis of the structure of precipitation patterns in New England. J. Appl. Meteorol. **11**, 926–935 (1972)

Ba, M.B., Gruber, A.: Goes multispectral rainfall algorithm (gmsra). J. Appl. Meteorol. **40**, 1500–1514 (2001)

Banerjee, S.: On geodetic distance computations in spatial modeling. Biometrics **61**(2), 617–625 (2005)

Banerjee, S., Gelfand, A.E., Knight, J.R., Sirmans, C.F.: Spatial modeling of house prices using normalized distance-weighted sums of stationary processes. J. Bus. Econ. Stat. **22**(2), 206–213 (2004)

Banerjee, S., Gelfand, A., Finley, A., Sang, H.: Gaussian predictive process models for large spatial data sets. J. R. Stat. Soc. B **70**(4), 825–848 (2008)

Bardossy, A., Plate, E.J.: Space-time model for daily rainfall using atmospheric circulation patterns. Water Resour. Res. **28**(5), 1247–1259 (1992)

Bell, T.L., Kundu, P.K.: A study of the sampling error in satellite rainfall estimates using optimal averaging of data and a stochastic model. J. Climate **9**, 1251–1268 (1996)

Bell, T.L., Abdullah, A., Martin, R.L., North, G.R.: Sampling errors for satellite-derived tropical rainfall: Monte Carlo study using a space-time stochastic model. J. Geophys. Res. **95**(D3), 2195–2205 (1990)

Bell, T.L., Kundu, P.K., Kummerow, C.D.: Sampling errors of ssm/i and trmm rainfall averages: comparison with error estimates from surface data and a simple model. J. Appl. Meteorol. **40**, 938–954 (2001)

Chakraborty, A., Gelfand, A.E., Wilson, A.M., Latimer, A.M., Silander, J.A.: Modeling large scale species abundance with latent spatial processes. Ann. Appl. Stat. **4**(3), 1403–1429 (2010)

Chakraborty, A., Mallick, B.K., McClarren, R.G., Kuranz, C.C., Bingham, D.R., Grosskopf, M.J., Rutter, E., Stripling, H.F., Drake, R.P.: Spline-based emulators for radiative shock experiments with measurement error. J. Am. Stat. Assoc. **108**, 411–428 (2013)

Chib, S., Carlin, B.P.: On mcmc sampling in hierarchical longitudinal models. Stat. Comput. **9**(1), 17–26 (1999)

Cohen, A.C.: Truncated and Censored Samples, 1st edn. Marcel Dekker, New York (1991)

Cooley, D., Nychka, D., Naveau, P.: Bayesian spatial modeling of extreme precipitation return levels. J. Am. Stat. Assoc. **102**(479), 824–840 (2007)

Cressie, N., Johannesson, G.: Fixed rank kriging for very large spatial data sets. J. R. Stat. Soc. B **70**(1), 209–226 (2008)

Denison, D.G.T., Mallick, B.K., Smith, A.F.M.: Bayesian mars. Stat. Comput. **8**(4), 337–346 (1998)

Felgate, D.G., Read, D.G.: Correlation analysis of the cellular structure of storms observed by raingauges. J. Hydrol. **24**, 191–200 (1975)

Finley, A., Sang, H., Banerjee, S., Gelfand, A.: Improving the performance of predictive process modeling for large datasets. Comput. Stat. Data Anal. **53**(8), 2873–2884 (2009)

Finley, A.O., Banerjee, S., MacFarlane, D.W.: A hierarchical model for quantifying forest variables over large heterogeneous landscapes with uncertain forest areas. J. Am. Stat. Assoc. **106**(493), 31–48 (2011)

Friedman, J.: Multivariate adaptive regression splines. Ann. Stat. **19**(1), 1–67 (1991)

Fuentes, M.: Spectral methods for nonstationary spatial processes. Biometrika **89**, 197–210 (2002)

Fuentes, M., Reich, B., Lee, G.: Spatial-temporal mesoscale modelling of rainfall intensity using gauge and radar data. Ann. Appl. Stat. **2**, 1148–1169 (2008)

Furrer, R., Genton, M.G., Nychka, D.: Covariance tapering for interpolation of large spatial datasets. J. Comput. Graph. Stat. **15**(3), 502–523 (2006)

Gelfand, A.E., Kim, H.J., Sirmans, C.F., Banerjee, S.: Spatial modeling with spatially varying coefficient processes. J. Am. Stat. Assoc. **98**(462), 387–396 (2003)

Gelfand, A.E., Banerjee, S., Finley, A.O.: Spatial design for knot selection in knot-based dimension reduction models. In: Mateu, J., Müller, W.G. (eds.) Spatio-Temporal Design: Advances in Efficient Data Acquisition, pp. 142–169. Wiley, Chichester (2012)

Guhaniyogi, R., Finley, A.O., Banerjee, S., Gelfand, A.E.: Adaptive Gaussian predictive process models for large spatial datasets. Environmetrics **22**(8), 997–1007 (2011)

Higdon, D.: A process-convolution approach to modelling temperatures in the North Atlantic Ocean. Environ. Ecol. Stat. **5**(2), 173–190 (1998)

Higdon, D.: Space and space-time modeling using process convolutions. In: Anderson, C., Barnett, V., Chatwin, P.C., El-Shaarawi, A.H. (eds.) Quantitative Methods for Current Environmental Issues, pp. 37–56. Springer, London (2002)

Higdon, D., Swall, J., Kern, J.: Non-stationary spatial modeling. In: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (eds.) Bayesian Statistics, vol. 7, pp. 181–197. Oxford University Press, Oxford (1999)

Huffman, G.J., Adler, R.F., Stocker, E.F., Bolvin, D.T., Nelkin, E.J.: A trmm-based system for real-time quasi-global merged precipitation estimates. In: TRMM International Science Conference, Honolulu, pp. 22–26 (2002)

Huffman, G.J., Adler, R.F., Bolvin, D.T., Gu, G., Nelkin, E.J., Bowman, K.P., Hong, Y., Stocker, E.F., Wolef, D.B.: The trmm multisatellite precipitation analysis (tmpa): quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. J. Hydrometeorol. **8**, 38–55 (2007)

Joyce, R.J., Janowiak, J.E., Arkin, P.A., Xie, P.: Cmorph: a method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. J. Hydrometeorol. **5**, 487–503 (2004)

Jun, M.: Non-stationary cross-covariance models for multivariate processes on a globe. Scand. J. Stat. **38**, 726–747 (2011)

Jun, M., Stein, M.L.: Nonstationary covariance models for global data. Ann. Appl. Stat. **2**(4), 1271–1289 (2008)

Kammann, E.E., Wand, M.P.: Geoadditive models. J. R. Stat. Soc., Ser. C, Appl. Stat. **52**(1), 1–18 (2003)

Kaufman, C.G., Schervish, M.J., Nychka, D.W.: Covariance tapering for likelihood-based estimation in large spatial data sets. J. Am. Stat. Assoc. **103**(484), 1545–1555 (2008)

Kidd, C.: Satellite rainfall climatology: a review. Int. J. Climatol. **21**, 1041–1066 (2001)

Lee, G.W., Zawadzki, I.: Variability of drop size distributions: time-scale dependence of the variability and its effects on rain estimation. J. Appl. Meteorol. **44**, 241–255 (2005)

Lethbridge, M.: Precipitation probability and satellite radiation data. Mon. Weather Rev. **95**(7), 487–490 (1967)

Marchenko, Y.V., Genton, M.G.: Multivariate log-skew-elliptical distributions with applications to precipitation data. Environmetrics **21**(3–4), 318–340 (2010)

McConnell, A., North, G.R.: Sampling errors in satellite estimates of tropical rain. J. Geophys. Res. **92**(D8), 9567–9570 (1987)

Negri, A.J., Xu, L., Adler, R.F.: A trmm-calibrated infrared rainfall algorithm applied over Brazil. J. Geophys. Res. **107**(D20), 8048–8062 (2002)

Paciorek, C., Schervish, M.: Spatial modelling using a new class of nonstationary covariance functions. Environmetrics **17**, 483–506 (2006)

Richardson, S., Green, P.J.: On Bayesian analysis of mixtures with an unknown number of components. J. R. Stat. Soc. B **59**(4), 731–792 (1997)

Rodríguez-Iturbe, I., Mejía, J.M.: The design of rainfall networks in time and space. Water Resour. Res. **10**, 713–728 (1974)

Sampson, P.D., Guttorp, P.: Nonparametric estimation on nonstationary spatial covariance structure. J. Am. Stat. Assoc. **87**, 108–119 (1992)

Sang, H., Gelfand, A.E.: Hierarchical modeling for extreme values observed over space and time. Environ. Ecol. Stat. **16**(3), 407–426 (2009)

Sang, H., Huang, J.Z.: A full scale approximation of covariance functions for large spatial data sets. J. R. Stat. Soc. B **74**(1), 111–132 (2012)

Schmidt, A.M., O'Hagan, A.: Bayesian inference for non-stationary spatial covariance structure via spatial deformations. J. R. Stat. Soc. B **65**, 743–758 (2003)

Simpson, J., Adler, R.F., North, G.R.: A proposed Tropical Rainfall Measuring Mission (TRMM) satellite. Bull. Am. Meteorol. Soc. **69**(3), 278–295 (1988)

Sorooshian, S., Hsu, K.L., Gao, X., Gupta, H., Imam, B., Braithwaite, D.: Evaluation of Persiann system satellite-based estimates of tropical rainfall. Bull. Am. Meteorol. Soc. **81**(9), 2035–2046 (2000)

Stein, M., Chi, Z., Welty, L.: Approximating likelihoods for large spatial data sets. J. R. Stat. Soc. B **66**, 275–296 (2004)

Sun, Y., Li, B., Genton, M.G.: Geostatistics for large datasets. In: Porcu, E., Montero, J.M., Schlather, M. (eds.) Advances and Challenges in Space-Time Modelling of Natural Events, vol. 207, pp. 55–77. Springer, Berlin (2012)

Tanner, T.A., Wong, W.H.: The calculation of posterior distributions by data augmentation. J. Am. Stat. Assoc. **82**, 528–549 (1987)

Vicente, G.A., Scofield, R.A., Menzel, W.P.: The operational goes infrared rainfall estimation technique. Bull. Am. Meteorol. Soc. **79**(9), 1883–1898 (1998)

Weng, F.W., Zhao, L., Ferraro, R., Pre, G., Li, X., Grody, N.C.: Advanced microwave sounding unit (amsu) cloud and precipitation algorithms. Radio Sci. **38**(4), 8068–8079 (2003)

Wilheit, T.T.: A satellite technique for quantitatively mapping rainfall rates over the ocean. J. Appl. Meteorol. **16**, 551–560 (1977)

Wilheit, T.T., Chang, A.T.C., Rao, M.S.V., Rodgers, E.B., Theon, J.S.: A satellite technique for quantitatively mapping rainfall rates over the oceans. J. Appl. Meteorol. **16**(5), 551–560 (1977)

Xie, P., Arkin, P.A.: Global monthly precipitation estimates from satellite-observed outgoing longwave radiation. J. Climate **11**, 137–164 (1998)