Journal of Environmental Statistics

February 2013, Volume 4, Issue 9. http://www.jenvstat.org

A Spatial Modelling Approach for the Blending and Error Characterization of Remotely Sensed Soil Moisture Products

Grace S. Chiu
CSIRO Mathematics, Informatics and Statistics

Eric A. Lehmann
CSIRO Mathematics, Informatics and Statistics

Joshua C. Bowden
CSIRO Information Management and Technology

Abstract

Soil moisture is one of the main physical quantities with a key role in water resources accounting research. Due to the spatial nature of this quantity and limitations of ground-based or remote sensing technology, the reliability of soil moisture data is of practical concern. Blending multiple sources of soil moisture data helps to combine the strengths and mitigate the weaknesses exhibited by each individual source. We build on the Bayesian hierarchical spatial model by Chiu and Lehmann (2011) to incorporate multiple sources of remotely sensed data on soil moisture. This model-based approach accounts for covariates, and can handle the various spatial resolutions among the data sources without manual aggregation or resampling. This unified approach also provides insights into the reliability (uncertainty) of each data source and of the blended product. We also briefly introduce an extension for model-based spatial aggregation of areal data.

Keywords: Bayesian inference, change-of-support problem, conditional autoregressive model, hierarchical modelling, Metropolis-Hastings, model-based spatial aggregation, remote sensing.

1. Introduction

According to the Australian Bureau of Meteorology, Australia is the driest inhabited continent. For reliable water resources assessment and accounting, continent-wide studies are required. Ground-based observations are considered most representative of the physical quantities relevant to this type of studies. Yet, they are spatially sparse and often lack geographical coverage to yield necessary data for large-scale investigations. Consequently, in the hydrological research community, remotely sensed products have played a vital role in water resources accounting. The term "product" refers to a physical quantity derived from surrogates (often via so-called retrieval models, e.g. Wigneron *et al.* 2003) instead of directly measured. Of the physical quantities relevant to water resources accounting research, soil moisture (SM) is one key component, as it directly influences surface runoff (Betson 1964). For example, the journal *Remote Sensing of Environment* alone publishes a vast number of articles that cover a wide range of topics related to SM, including instrumentation technology (e.g. Aubert *et al.* 2011); retrieval models and their performance for converting captured surrogate raster images (of, say, surface temperature) to SM (e.g. Wigneron *et al.* 2003); benchmarking of remotely sensed SM products against deterministic model-based products (e.g. Doubková *et al.* 2012); and many more. Other journals with a heavy presence of articles on remotely sensed SM products include the *IEEE Transactions on Geoscience and Remote Sensing*.

At the start of October 2011, major instrument failure occurred to the Advanced Microwave Scanning Radiometer – Earth Observing System (AMSR-E) aboard the Aqua satellite. Until its failure, the AMSR-E SM product had been regarded as one of the more reliable sources of areal SM data with wide coverage. Thus, much remote sensing literature on SM has featured AMSR-E. Other frequently studied satellite-based SM products include those derived from images captured by the Advanced Synthetic Aperture Radar (ASAR) and the Advanced Scatterometer (ASCAT), aboard the Envisat and Metop-A satellites, respectively.

Areal products are available not only by remote sensing, but also by deterministic modelling, including the CSIRO Atmosphere Biosphere Land Exchange (CABLE) (Kowalczyk *et al.* 2006) and the Australian Water Resources Assessment Landscape (AWRA-L) (van Dijk and Warren 2010; van Dijk and Renzullo 2011) models. When multiple areal SM products are available, making the best collective use of such information sources is of obvious interest. The blending/fusion/synthesis of multiple products to produce a single hybrid product is one of the most active areas of water resources accounting research.

Typically, different remote sensing instruments record emitted or reflected radiation at different pixel sizes; the readings are then converted to estimates of biogeochemical quantities, such as SM, using a range of empirical and analytical models. Products derived from deterministic models may come at their own spatial resolutions, too. Pixel-wise blending after manual spatial realignment is a popular approach for blending areal products. In this approach, a working resolution is prescribed from a candidate set of one or more resolutions. Point-level or areal products with pixel sizes smaller than prescribed are spatially aggregated (upscaled) by some form of averaging, and conversely, products with larger pixel sizes are "re-sampled" or "sub-sampled" (downscaled) to the working resolution, typically via some interpolation algorithm. Correlation among the reprocessed raster images is assessed for each candidate resolution, and the apparently optimal candidate is chosen. Within a given pixel at the chosen resolution, blending of SM products is peformed separately from other pixels, then individual pixels are stitched back together to form a raster image of the blended product. Researchers in the AWRA Model-Data Fusion (MDF) Project (under the Water Information Research and Development Alliance between CSIRO and the Australian Bureau of Meteorology), e.g. Gouweleeuw *et al.* (2011), Jin and Henderson (2011), and Renzullo *et al.* (2012), have built on existing work from the literature to demonstrate different pixel-wise blending approaches.

To date, we are unaware of articles prior to Chiu and Lehmann (2011) that account for within-product spatial dependence for blending multiple products in water resources moni-

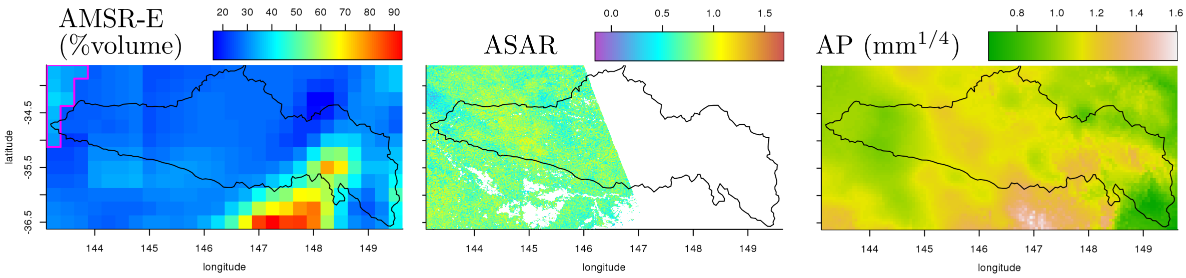

Figure 1: MRC quantities on 2005-06-13. AMSR-E data are missing over pixels outlined in pink, and values shown for those pixels are imputed through the posterior predictive distribution based on our model (see Sections 3 and 5.2). White ASAR pixels denote "blind spots" where data are missing/irrelevant (see Sections 2 and 3.2). Antecedent precipitation (AP) is a transformation of AWAP (see Section 3.3).

toring research. These authors, also in the AWRA MDF Project, propose a Bayesian hierarchical framework with conditional autoregressive (CAR) spatial random effects to model the relationship among two SM products and two covariates; the hierarchical structure handles the change-of-support problem and avoids manual spatial realignment. The unified framework not only produces a blended SM product, but yields rigorous uncertainty estimates and allows benchmarking of one against the other, all in a single data analysis. On the other hand, their main objective was not a blended product *per se*, as one of the products was ground based and contributed minimally to the blended product. In this paper, we build on their approach to blend two remotely sensed SM products alongside covariates. We also demonstrate statistical model-based error characterization for these products through our framework. Chiu and Lehmann (2011) considered the rectangular region that circumscribes the Murrumbidgee River Catchment in eastern Australia (outlined in black, e.g. in Fig. 1), due to a lack of ground SM measurements beyond the catchment for benchmarking. Although we do not model ground measurements in this paper, we consider the same rectangular region so that the practicality of our methodology can be demonstrated for a manageable portion of the continent before it is extended to continental scale. We focus on daily data from January and June to represent typical summer and winter conditions, respectively. Although temporal smoothing is absent from our approach, an integrated spatio-temporal model (e.g. Cressie and Wikle 2011) may be pursued to additionally address temporal patterns.

In Section 2, we describe the two SM products to be blended, and one of the covariates (the remaining covariate is spatial coordinate). Our hierarchical model is presented in Section 3. The Bayesian inference and implementation procedure is described in Section 4, followed by inference results in Section 5, including error characterization of the input and blended products. We conclude our paper in Section 6, in which we also describe an extension for model-based spatial aggregation of a given product, and offer directions of future extension.

2. Sources of data

For our hierarchical spatial model in Section 3, we consider the following MRC quantities.

The AWAP product for daily rainfall. The Australian Water Availability Project (AWAP) precipitation product (Jones *et al.* 2009) in mm is gridded over 0.05° pixels. It results from applying smoothing spline interpolation to rainfall "anomalies" defined using measurements from a spatial network of rain gauges. We consider AWAP data to coincide with the time frame of remote sensing data described below. There are no missing AWAP values across Australia. We consider AWAP data for the MRC only.

The remotely-sensed VUA AMSR-E product for soil moisture. We consider the AMSR-E SM product by Vrije Universiteit Amterdam (Owe *et al.* 2008). AMSR-E data over 0.25° pixels are available for the entire Australian mainland except for "blind spots" (due to a combination of heavy vegetation, unfavourable weather conditions, instrumentation failure, etc.). The AMSR-E product is recorded in %volume; it is regarded as a direct representation of SM. Edges of AMSR-E pixels align perfectly with AWAP pixel edges. Thus, each AMSR-E pixel contains exactly 25 AWAP pixels. For spatial modelling, we consider C-band data from the descending pass over the MRC only, comprsing M=312 AMSR-E pixels.

Daily ASAR wetness index with masked pixels. In the Global Monitoring mode, ASAR records backscatter images around the globe in swaths of 30 arc-second pixels. Partial coverage of Australia is through up to several swaths per day. We use the prototype ASAR surface wetness product developed by the Technische Universität Wien (TUW) and provided to CSIRO for assessing a deterministic-model-based SM product (van Dijk and Warren 2010; Doubková *et al.* 2012). This prototype is a reformulation of original images into 15 arc-second pixels (i.e. 144 such pixels nested in each AWAP pixel) by correcting for the incidence angle at which the backscatter instrument operates. Our ASAR time frame essentially spans 2005 to 2008. Certain ASAR pixels are masked, i.e. pre-identified as locations where this ASAR product may be uninformative. We treat masked pixels as non-existent throughout the time frame. The natural range of the ASAR product is [0, 1], and is interpreted as a unitless index of relative wetness rather than a direct representation of SM (van Dijk and Warren 2010). The definition of ASAR relative wetness requires estimating driest and wettest conditions (Mladenova *et al.* 2010) and can produce ASAR values outside of [0, 1]. Rather than treating out-of-range ASAR values as missing data, we adopt the common practice in hydrology of regarding them as indications of extreme dryness or wetness relative to the full dataset. Indeed, according to histograms[1] of ASAR values from selected pixels (approximately equidistant) across the MRC, the empirical distribution of ASAR values is smooth and roughly Gaussian, with no indication of anomaly due to out-of-range values.

3. A Hierarchical Statistical Spatial Model

Fig. 2 provides a visual representation of the hierarchical structure among key quantities in our statistical model. We consider AWAP to be the underlying driver of latent SM. Thus, latent SM is modelled at the same spatial resolution as AWAP. To this end, we adapt the model by Chiu and Lehmann (2011), with ground probe data replaced by ASAR data. The rationale is as follows. For Chiu and Lehmann (2011), one of the objectives was to benchmark the AMSR-E data (areal, contiguous over the continent except for blind spots) relative to ground probe data (point-level, very sparse, and available only for 38 MRC locations). For this purpose, the contribution of the latter towards the inference for latent SM (areal, contiguous over

[1]Not shown. They comprise ASAR values from selected pixels altogether and temporally collapsed either over January from all of 2005 to 2009, or over June from all of 2005 to 2008.

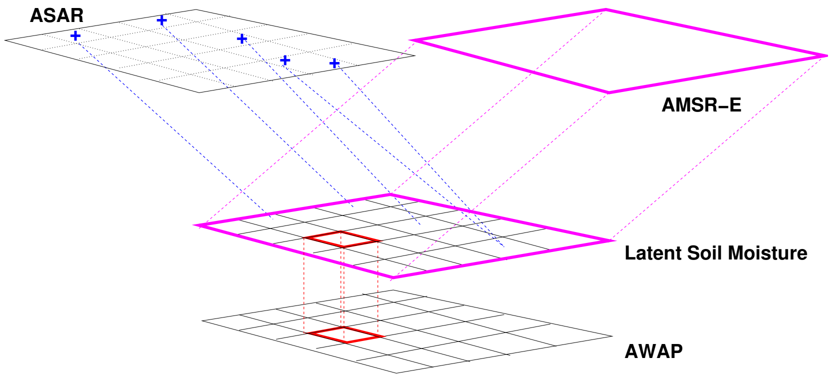

Figure 2: Schematic for the relationship among AMSR-E, ASAR, AWAP (observable) and SM (latent) within the model hierarchy. Pink outlines delimit an AMSR-E pixel, and red outlines, an AWAP or SM pixel. ASAR pixels (very small) are represented by blue "+" signs. Each set of dotted lines joining two consecutive levels in the hierarchy corresponds to a regression layer in the model. Specifically, ASAR is a response of SM at the SM resolution; AMSR-E is a response of SM at the AMSR-E resolution; and SM is a response of AWAP at a common resolution. Spatial autocorrelation at each level is attributable to that among AWAP pixels.

the MRC) is understandably minimal. In contrast, currently our objective is the inference for latent SM and the error characterization of AMSR-E; thus, ASAR data (areal at a very high resolution, available over the continent except for blind spots) is expected to contribute substantively towards this inference when compared to ground probe data. Ignoring ground probe data altogether in the current model removes the burden of extra model parameters.

Despite some concern over ASAR's ability to inform the underlying SM content (Mladenova *et al.* 2010; Pathe *et al.* 2009), the role of the ASAR wetness index in the spatial model hierarchy may be regarded as ancillary information to be "assimilated" into the AMSR-E product, the latter being a gridded representation of SM. Both products are in turn informed by the AWAP product, whose spatial variation patterns are passed along to the latent state and further to both ASAR and AMSR-E. Another argument for including potentially less informative data is to safeguard against "biased assimilation": some investigators are proned to consider only data that agree with their world view while ignoring all other data; this practice often causes biased conclusions (Lord and Taylor 2009).

Our model statements below are intended to handle the spatial structure of MRC quantities for a given day.

3.1. Relating AMSR-E data and big-pixel SM state

Let B_m be the mth AMSR-E pixel and s_{mr} the rth AWAP pixel inside B_m, for r=1, …, 25 and m=1, …, M; $v_m \equiv v(B_m)$ the AMSR-E value at B_m; $\psi_{mr} \equiv \psi(s_{mr})$ the latent SM spatially averaged across s_{mr}; and $\overline{\psi}_m \equiv \psi(B_m) = (1/25)\sum_r \psi_{mr}$ the latent SM spatially averaged across B_m. Assuming AMSR-E to be a change-of-scale (i.e. linear) representation of latent SM, then

$$v_m = \alpha_0 + \alpha_1 \overline{\psi}_m + \delta_m\,, \qquad \delta_m \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma_\delta^2)\,. \tag{1}$$

3.2. Relating ASAR data and SM state inside an AWAP pixel

Let q_{mrk} be the ASAR value at the kth observed pixel inside s_{mr} for k=1, ..., K_{mr} (defined only for AWAP pixels with $K_{mr}{\geq}1$); and $\tilde{\psi}_{mr} = \log\,\psi_{mr}$. Then, we take

$$q_{mrk} = \alpha_3\tilde{\psi}_{mr} + \varepsilon_{mrk}, \quad \varepsilon_{mrk} \overset{\text{ind}}{\sim} \mathrm{N}(0, \sigma_m^2), \quad (m, r, k) \in \mathcal{M}_q \equiv \{(m, r, k) : q_{mrk} \text{ observed}\}. \tag{2}$$

Pairwise scatterplots (not shown) suggest approximate linearity between log(AMSR-E) and ASAR, so that Eq. (1) implies log-transformation of ψ in (2). Again, we regard q_{mrk} as point-level data due to the high resolution of ASAR images and the abundance of ASAR blind spots which include masked pixels. On any given day, ASAR blind spots comprise vast contiguous regions of the MRC (e.g. Fig. 1, middle panel). Thus, instead of imputing q at blind spots, we consider each captured ASAR image to correspond to fixed spatial design points.[2] Finally, the lack of an intercept in (2) is to avoid potential confounding with β_0 in (3) below.

3.3. Relating SM state and antecedent precipitation inside an AWAP pixel

We wish to model the latent SM as being driven by antecedent precipitation (AP), subject to additive error with constant variance σ_η^2. Then,

$$\tilde{\psi}_{mr} = \beta_0 + \beta_1(p_{mr} - p^*) + \eta_{mr}\,, \qquad \eta_{mr} \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma_\eta^2)\,, \tag{3}$$

where $p_{mr}{\equiv}p(s_{mr})$ is the 4th root of a 20-day moving average of AWAP at s_{mr} (see Chiu and Lehmann 2011, for rationale), and p^* is an arbitrary pre-specified constant; taking p^* as approximately the mean of the observed $\{p_{mr}\}$ is known as covariate centring, which reduces the dependence between β_0 and β_1, thus improving MCMC mixing (e.g. see Gelman and Hill 2007). According to further exploratory analyses, our definition of p_{mr} and log-transformation of ψ (based on log(AMSR-E)) in (3) is appropriate for June, yet less so for January. Due to a noticeable difference in relationships among relevant quantities between these months, future extensions of our model can include a separate summer structure for the p-ψ relationship. In this paper, we assume (3) for both months.

3.4. Relating antecedent precipitation and space

Let $\boldsymbol{x}_{mr}{=}(x_{mr1}, x_{mr2})$ be a linear transformation of the longitude-latitude coordinates of the top-left corner of s_{mr}; $h_{\boldsymbol{\gamma}}$ a polynomial of $\boldsymbol{x}_{mr}$ with a pre-specified order and coefficient vector $\boldsymbol{\gamma}$; and $\phi_{mr}{\equiv}\phi(s_{mr})$ the spatial random error associated with s_{mr}. Then,

$$p_{mr} = h_{\boldsymbol{\gamma}}(\boldsymbol{x}_{mr}) + \phi_{mr} + \zeta_{mr}\,, \qquad \zeta_{mr} \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma_\zeta^2)\,. \tag{4}$$

We take $h_{\boldsymbol{\gamma}}(\boldsymbol{x}_{mr}){=}\gamma_1x_{mr1}{+}\gamma_2x_{mr2}{+}\gamma_3x_{mr1}x_{mr2}{+}\gamma_4x_{mr1}^2{+}\gamma_5x_{mr2}^2$, as various earlier models using a higher order polynomial with interactions suggested that this formulation of h suffices.

3.5. Spatial patterns

Reindex $\{\phi_{mr}\}$ as $\boldsymbol{\phi}{=}[\phi_1, \phi_2, \ldots, \phi_{25M}]'$. We take a special nearest-neighbour Gaussian CAR model for $\boldsymbol{\phi}$, denoted by

$$\boldsymbol{\phi} \sim \mathrm{CAR}(1; \tau^2) \tag{5}$$

[2] At the cost of higher computational intensity, the current model may be modified to regard ASAR data as areal, treating blind spots and out-of-range values as missing data to be imputed.

with scale parameter τ^2 and "1" corresponding to a first-order rectangular adjacency of AWAP pixels that induces spatial dependence.[3] Specifically, let $\boldsymbol{\phi}_{-i}$ denote $\boldsymbol{\phi}$ with ϕ_i removed. Then,

$$\phi_i \left| \boldsymbol{\phi}_{-i}, \tau^2 \right. \sim \mathrm{N}\left(\frac{1}{w_{i+}} \sum_{i' \neq i} w_{ii'} \phi_{i'}, \frac{\tau^2}{w_{i+}} \right)$$

where the dependence structure of $\boldsymbol{\phi}$ is imposed by

$$w_{ii'} = \begin{cases} 1 & \text{if } s_{i'} \text{ is one of the 8 neighbours of } s_i \text{ along its rectangular border} \\ 0 & \text{otherwise} \end{cases}$$

with $w_{i+}{=}\sum_{i'=1}^{25M} w_{ii'}$. Note that $w_{ii}{=}0$ for all i. Taking $\sum_i \phi_i{=}0$, the log-likelihood of $\boldsymbol{\phi}$ given τ^2 is (Banerjee *et al.* 2004)

$$\log f(\boldsymbol{\phi}|\tau^2) = -\frac{25M-1}{2} \log \tau^2 - \frac{1}{2\tau^2} \boldsymbol{\phi}'(\mathbb{D} - \mathbb{W})\boldsymbol{\phi} + \text{constant}$$

where $\mathbb{W}{=}[w_{ii'}]$ is the symmetric adjacency matrix for our model, and $\mathbb{D}{=}\mathrm{diag}\{w_{1+}, \ldots, w_{25M+}\}$. Note that $\mathbb{D} - \mathbb{W}$ has rank $25M{-}1$.

3.6. Prior distributions

Finally, we take

$$\alpha_1, \alpha_3, \beta_1 \overset{\text{iid}}{\sim} \mathrm{N}(3a^{-1/2}, a^{-1}), \quad \alpha_0, \beta_0, \gamma_\ell \overset{\text{iid}}{\sim} \mathrm{N}(0, a^{-1}), \quad \sigma_m^2, \sigma_\delta^2, \sigma_\eta^2, \sigma_\zeta^2, \tau^2 \overset{\text{ind}}{\sim} \mathrm{IG}(a_1, a_2) \tag{6}$$

for all $\ell{=}1, \ldots, 5$ and all $m{\in}\mathcal{S}{\equiv}\{m\colon \sum_r K_{mr}{\geq}\ 1\}$. Here, $\mathrm{IG}(a_1, a_2)$ is the inverse-Gamma distribution with mode $a_2/(a_1{+}1)$. Centring the slope priors at three times the prior standard deviation (SD) is to impose a large prior probability of positive association between q and $\tilde{\psi}$, between v and ψ, and between $\tilde{\psi}$ and p. For convenience and relative diffuseness, we take $a = 0.01$, and $a_1 = a_2 = 1$.

4. Bayesian inference

Let $\boldsymbol{v}^{\text{obs}}$ be the vector of $\{v_m\}$ where observed. Also let $\boldsymbol{v}^{\text{mis}}$ be the vector of $\{v_m\}$ where unobserved. Then, the posterior distribution for our model parameters is

$$\begin{aligned} & f\left(\boldsymbol{v}^{\text{mis}}, \boldsymbol{\psi}, \boldsymbol{\phi}, \boldsymbol{\Omega} \middle| \, \text{data}\right) = f\left(\boldsymbol{v}^{\text{mis}}, \boldsymbol{\psi}, \boldsymbol{\phi}, \boldsymbol{\Omega} \middle| \, \boldsymbol{q}, \boldsymbol{v}^{\text{obs}}, \boldsymbol{p}\right) \\ \propto \quad & f(\boldsymbol{q}, \boldsymbol{v}, \boldsymbol{p}, \boldsymbol{\psi}, \boldsymbol{\phi}, \boldsymbol{\Omega}) \\ = \quad & f\left(\boldsymbol{q} \middle| \, \alpha_3, \boldsymbol{\psi}, \{\sigma_m^2\}\right) f\left(\boldsymbol{v} \middle| \, \alpha_0, \alpha_1, \boldsymbol{\psi}, \sigma_\delta^2\right) f\left(\boldsymbol{\psi} \middle| \, \boldsymbol{\beta}, \boldsymbol{p}, \sigma_\eta^2\right) f\left(\boldsymbol{p} \middle| \, \boldsymbol{\phi}, \boldsymbol{\gamma}, \sigma_\zeta^2\right) f\left(\boldsymbol{\phi} \middle| \, \tau^2\right) f(\boldsymbol{\Omega}) \end{aligned} \tag{7}$$

where $\boldsymbol{\Omega}{=}[\boldsymbol{\alpha}', \boldsymbol{\beta}', \boldsymbol{\gamma}', \{\sigma_m^2 : m = 1, \ldots, M\}, \sigma_\delta^2, \sigma_\eta^2, \sigma_\zeta^2, \tau^2]'$, and $f(\boldsymbol{\Omega})$ is the prior from (6). The posterior (7) has no closed form; we approximate it via Markov chain Monte Carlo (MCMC),

[3] When the 5th-order exponential decay version of (5) by Chiu and Lehmann (2011) was applied to blend AMSR-E and ASAR, the overall inference showed minimally noticeable difference to the first-order model here. Moreover, using positive weights for all areal neighbours in a higher-order CAR model may be unncessarily restrictive; for example, see Lindgren *et al.* (2011).

from which inference for the unknown quantities $\boldsymbol{v}^{\text{mis}}, \boldsymbol{\psi}, \boldsymbol{\phi}$, and $\boldsymbol{\Omega}$ is drawn. Sparse matrix computational algorithms allow major reduction in computational time, while parallelization within MCMC iterations can lead to further reduction (see Appendix A1).

4.1. Full conditionals

The numbered equations below are full conditional distributions for MCMC implementation of (7).

$$f\left(\boldsymbol{v}^{\text{mis}}|\,\boldsymbol{.}\right) = \prod_{m:v_m \text{ unobserved}} \mathrm{N}\left(\alpha_0 + \alpha_1\overline{\psi}_m, \sigma_\delta^2\right)\,, \tag{8}$$

$$\log f(\psi_{mr}|\boldsymbol{.}) = -\frac{1}{2}\{c_{11mr}\log\psi_{mr} + c_{21mr}(\log\psi_{mr})^2 + c_{12mr}\psi_{mr} + c_{22mr}\psi_{mr}^2\} + \text{constant} \tag{9}$$

$$c_{11mr} = 2\left[1 - \frac{\alpha_3\sum_k q_{mrk}}{\sigma_m^2} - \frac{\beta_0 + \beta_1(p_{mr}-p^*)}{\sigma_\eta^2}\right], \qquad c_{21mr} = \frac{\alpha_3^2 K_{mr}}{\sigma_m^2} + \frac{1}{\sigma_\eta^2}\,,$$

$$c_{12mr} = \frac{-2\alpha_1}{25\sigma_\delta^2}\left(v_m - \alpha_0 - \frac{\alpha_1}{25}\sum_{r'\neq r}\psi_{mr'}\right), \qquad c_{22mr} = \frac{\alpha_1^2}{25^2\sigma_\delta^2}\,,$$

$$\phi_i|\boldsymbol{.} \sim \mathrm{N}\left(\frac{c_{01i}}{c_{02i}}, \frac{1}{c_{02i}}\right) \qquad \text{with constraint } \sum_i \phi_i = 0 \tag{10}$$

$$c_{01i} = \frac{p_i - h_{\boldsymbol{\gamma}}(\boldsymbol{x}_i)}{\sigma_\zeta^2} + \frac{1}{\tau^2}\sum_{i'} w_{ii'}\phi_{i'}\,, \qquad c_{02i} = \frac{1}{\sigma_\zeta^2} + \frac{w_{i+}}{\tau^2}\,,$$

$$\alpha_3|\boldsymbol{.} \sim \mathrm{N}\left(\frac{c_1}{c_2}, \frac{1}{c_2}\right), \quad c_1 = \sum_{(m,r,k)\in\mathcal{M}_q} \frac{\tilde{\psi}_{mr}q_{mrk}}{\sigma_m^2} + 3\sqrt{a}\,, \quad c_2 = \sum_{m,r}\frac{K_{mr}\tilde{\psi}_{mr}^2}{\sigma_m^2} + a\,, \tag{11}$$

$$[\alpha_0, \alpha_1]'|\boldsymbol{.} \sim \mathrm{BVN}\left(\mathbb{C}_8^{-1}\boldsymbol{c}_7, \mathbb{C}_8^{-1}\right) \tag{12}$$

$$\boldsymbol{c}_7 = \sigma_\delta^{-2}\,\boldsymbol{\Psi}'\boldsymbol{v} + a\boldsymbol{\mu}\,, \quad \mathbb{C}_8 = \sigma_\delta^{-2}\,\boldsymbol{\Psi}'\,\boldsymbol{\Psi} + a\mathbb{I}_2\,,$$

$$\boldsymbol{\mu} = \begin{bmatrix} 0 \\ \frac{3}{\sqrt{a}} \end{bmatrix}, \qquad \boldsymbol{\Psi} = \begin{bmatrix} 1 & \overline{\psi}_1 \\ \vdots & \vdots \\ 1 & \overline{\psi}_M \end{bmatrix}, \qquad \mathbb{I}_n = n \times n \text{ identity matrix,}$$

$$\boldsymbol{\beta}|\boldsymbol{.} \sim \mathrm{BVN}\left(\mathbb{C}_4^{-1}\boldsymbol{c}_3, \mathbb{C}_4^{-1}\right) \tag{13}$$

$$\boldsymbol{c}_3 = \sigma_\eta^{-2}\mathbb{P}'\tilde{\boldsymbol{\psi}} + a\boldsymbol{\mu}\,, \qquad \mathbb{C}_4 = \sigma_\eta^{-2}\mathbb{P}'\mathbb{P} + a\mathbb{I}_2\,,$$

$$\tilde{\boldsymbol{\psi}}_m = \begin{bmatrix} \tilde{\psi}_{m,1} \\ \vdots \\ \tilde{\psi}_{m,25} \end{bmatrix}, \qquad \tilde{\boldsymbol{\psi}} = \begin{bmatrix} \tilde{\boldsymbol{\psi}}_1 \\ \vdots \\ \tilde{\boldsymbol{\psi}}_M \end{bmatrix}, \qquad \mathbb{P}_m = \begin{bmatrix} 1 & p_{m,1} - p^* \\ \vdots & \vdots \\ 1 & p_{m,25} - p^* \end{bmatrix}, \qquad \mathbb{P} = \begin{bmatrix} \mathbb{P}_1 \\ \vdots \\ \mathbb{P}_M \end{bmatrix},$$

$$\boldsymbol{\gamma}|\boldsymbol{.} \sim \mathrm{MVN}\left(\mathbb{C}_6^{-1}\boldsymbol{c}_5, \mathbb{C}_6^{-1}\right) \tag{14}$$

$$\boldsymbol{c}_5 = \frac{\mathbb{H}'(\boldsymbol{p} - \boldsymbol{\phi})}{\sigma_\zeta^2}, \; \mathbb{C}_6 = \frac{\mathbb{H}'\mathbb{H}}{\sigma_\zeta^2} + a\mathbb{I}_5\,, \boldsymbol{h}_{mr} = \begin{bmatrix} x_{mr1} \\ x_{mr2} \\ x_{mr1}x_{mr2} \\ x_{mr1}^2 \\ x_{mr2}^2 \end{bmatrix}, \; \mathbb{H}_m = \begin{bmatrix} \boldsymbol{h}'_{m,1} \\ \vdots \\ \boldsymbol{h}'_{m,25} \end{bmatrix}, \; \mathbb{H} = \begin{bmatrix} \mathbb{H}_1 \\ \vdots \\ \mathbb{H}_M \end{bmatrix},$$

$$\tau^2\big|\,\boldsymbol{.} \sim \mathrm{IG}\left(a_1 + (25M-1)/2, a_2 + (1/2)\boldsymbol{\phi}'(\mathbb{D} - \mathbb{W})\boldsymbol{\phi}\right)\,, \tag{15}$$

$$\sigma_m^2 \big| \boldsymbol{\cdot} \sim \mathrm{IG}\Big(a_1 + (1/2)\sum_r K_{mr}, a_2 + (1/2)\sum_r \sum_{k=1}^{K_{mr}} (q_{mrk} - \alpha_3 \tilde{\psi}_{mr})^2\Big), \tag{16}$$

$$\sigma_\delta^2 \big| \boldsymbol{\cdot} \sim \mathrm{IG}\Big(a_1 + M/2, a_2 + (1/2)\sum_m (v_m - \alpha_0 - \alpha_1 \overline{\psi}_m)^2\Big), \tag{17}$$

$$\sigma_\eta^2 \big| \boldsymbol{\cdot} \sim \mathrm{IG}\Big(a_1 + 25M/2, a_2 + (1/2)\sum_{m,r} [\tilde{\psi}_{mr} - \beta_0 - \beta_1(p_{mr} - p^*)]^2\Big), \tag{18}$$

$$\sigma_\zeta^2 \big| \boldsymbol{\cdot} \sim \mathrm{IG}\Big(a_1 + 25M/2, a_2 + (1/2)\sum_{m,r} \left(p_{mr} - \boldsymbol{h}'_{mr}\boldsymbol{\gamma} - \phi_{mr}\right)^2\Big). \tag{19}$$

4.2. Metropolis-Hastings step for updating ψ_{mr}

Unlike Chiu and Lehmann (2011), here we cannot implement (8)–(19) with a pure Gibbs sampler due to the non-standard distribution of ψ_{mr} from (9). Instead, we employ a Metropolis-Hastings (MH) step (e.g. see Hoff 2009) for (9). First, take

$$J(\psi_{mr}|z,\nu) = \frac{\nu^\nu}{\Gamma(\nu)} \frac{\psi_{mr}^{\nu-1}\exp\{-\nu\psi_{mr}/z\}}{z^\nu} \tag{20}$$

to be the proposal distribution for ψ_{mr}. This is a Gamma distribution (positive real) with mean z and shape parameter ν (so that the scale parameter is z/ν). Here, ν is the MH tuning parameter chosen so that J is right-skewed and unimodal.[4] Next, the acceptance ratio is

$$R_\nu\left(\psi^*_{mr}, \psi^{(t)}_{mr}\right) = \frac{f\left(\psi^*_{mr} \,\middle|\, \boldsymbol{\theta}^{(t)}_{-\psi_{mr}}, \boldsymbol{q}, \boldsymbol{v}^{\mathrm{obs}}, \boldsymbol{p}\right)}{f\left(\psi^{(t)}_{mr} \,\middle|\, \boldsymbol{\theta}^{(t)}_{-\psi_{mr}}, \boldsymbol{q}, \boldsymbol{v}^{\mathrm{obs}}, \boldsymbol{p}\right)} \times \frac{J\left(\psi^{(t)}_{mr} \,\middle|\, \psi^*_{mr}, \nu\right)}{J\left(\psi^*_{mr} \,\middle|\, \psi^{(t)}_{mr}, \nu\right)} \implies \tag{21}$$

$$\log R_\nu\left(\psi^*_{mr}, \psi^{(t)}_{mr}\right) = \left(\frac{c_{11mr}}{2} + 2\nu - 1\right)\log\frac{\psi^{(t)}_{mr}}{\psi^*_{mr}} + \frac{c_{21mr}}{2}\left[(\log\psi^{(t)}_{mr})^2 - (\log\psi^*_{mr})^2\right] + \tag{22}$$
$$\frac{c_{12mr}}{2}\left(\psi^{(t)}_{mr} - \psi^*_{mr}\right) + \frac{c_{22mr}}{2}\left[(\psi^{(t)}_{mr})^2 - (\psi^*_{mr})^2\right] + \nu\left(\frac{\psi^*_{mr}}{\psi^{(t)}_{mr}} - \frac{\psi^{(t)}_{mr}}{\psi^*_{mr}}\right)$$

where f and J in (21) are from (9) and (20), respectively.

5. Inference results

As it was infeasbile to apply our current modelling framework and computer implementation to all days for which AMSR-E, ASAR, and AP data are available, we focused on selected January and June dates between 2005 and 2009 to demonstrate our methodology. Selection criteria were largely based on data abundance for both SM products within the MRC boundaries. This was because (a) large amounts of missing AMSR-E data would substantially increase the already high-dimensional parameter space and thus weaken the inference, and (b) blending of the two SM products would be less meaningful given a small swath of ASAR image. For a few of the selected dates, MCMC mixing was such that convergence could not

[4] For our data, MCMC mixing is satisfactory for many datasets with ν=2. Our other values of ν led to little difference in mixing.

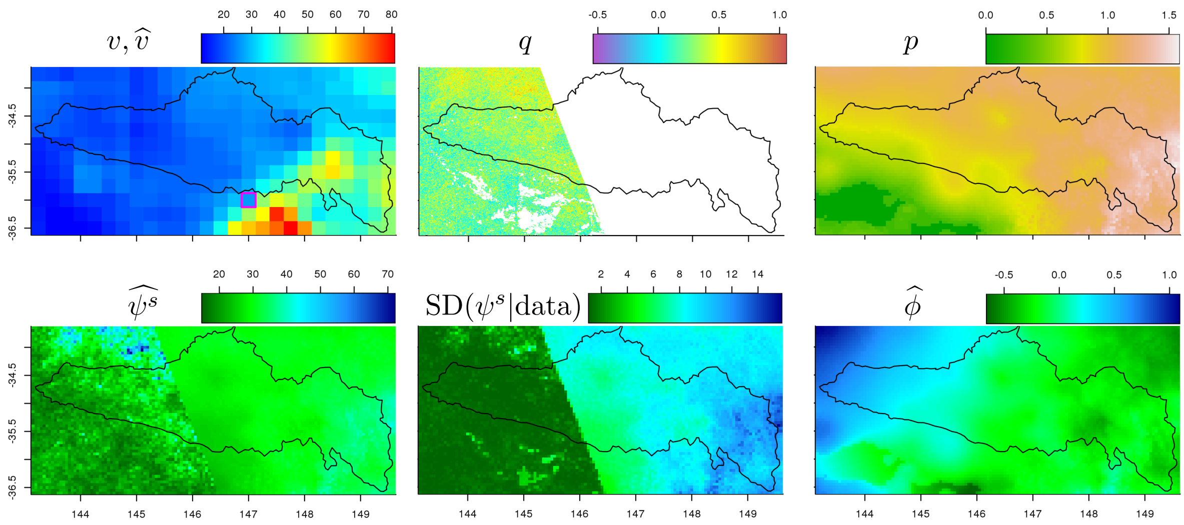

Figure 3: MRC data and inference for 2006-06-17. A "$\widehat{\ }$" denotes the posterior mean as an estimate of the corresponding quantity, e.g. $\widehat{v}$ refers to imputed AMSR-E (pixel outlined in pink).

be reached within a reasonable run-time, although this might be mitigated by incorporating temporal smoothing via an integrated spatio-temporal model. While such extension is intended for the future, in this paper we report results on the remaining eight selected dates that yielded proper Bayesian inference for SM over the MRC under manageable run-times.

5.1. Latent soil moisture — a blended product

Note that according to the model in Section 3, ψ is merely a quantity that is driven by AP (p), which in turn is the driver for AMSR-E (v) and ASAR (q) observations. For communications purposes, we define the scaled latent SM, $\psi^s_{mr} = \alpha_0 + \alpha_1 \psi_{mr}$, so that according to (1), both ψ^s_{mr} and v are in %volume; the former is the quantity of main interest when referring to the inference for SM. Specifically, $\widehat{\psi^s}$ (posterior mean of ψ^s) can be considered an SM product from blending AMSR-E and ASAR while making use of covariate information in the form of AP and spatial coordinates.

Fig. 3 presents some data and inference maps for 2006-06-17. This figure (and all others in Supp. Section S1) shows that the blended product $\widehat{\psi^s}$ and its uncertainty (represented by the posterior SD) are not spatially smooth, clearly showing the footprint of ASAR due to its limited coverage of the MRC. This feature is not surprising, as the model accounts for ASAR as well as AMSR-E only when data are present for both. Over the large regions where ASAR data are absent, the inference for ψ^s is due only to AMSR-E, if observed, alongside covariates. Thus, unless all the AMSR-E, ASAR, and AP datasets are mutually consistent, one can expect the ASAR footprint to be visible in $\widehat{\psi^s}$ and SD(ψ^s|data) — for the latter, in the form of low uncertainty relative to the rest of the MRC. The ASAR footprint is understandably more pronounced when the three datasets are less consistent, e.g. 2006-06-17.

Aside from latent SM, the spatial random effect, ϕ, is of additional interest, at least from a modelling perspective. Within the model hierarchy, ϕ accounts for spatial patterns of AP not addressed by the spatial coordinates through the trend surface $h_{\boldsymbol{\gamma}}(\,)$. With the coefficients $\boldsymbol{\gamma}$

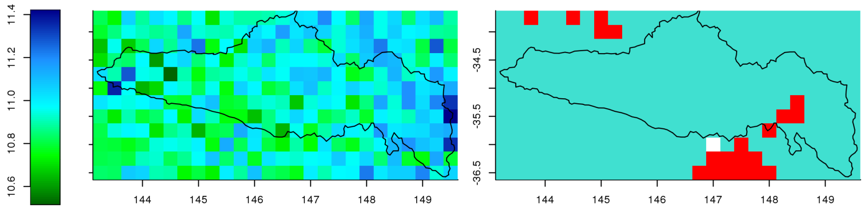

Figure 4: AMSR-E posterior predictive inference for 2006-06-17: SD (left, green-blue colour spectrum); coverage of nominal 95% CIs (right, tritone) with turquoise=covered, red=missed, white=no data.

being non-negligible,[5] clear spatial features of $\widehat{\phi}$ would suggest the crucial role that it plays in the model hierarchy. Such features are obvious in Fig. 3; this is indeed the case for all dates examined (see Supp. Section S1 for other dates).

5.2. Error characterization for AMSR-E and in-sample model validation

To evaluate the model-based uncertainty of AMSR-E data across the MRC under the statistical framework of this paper, we examine $f(\boldsymbol{v}^{\text{pred}}|\text{data})$, the posterior predictive distribution for AMSR-E (see Appendix A2). (Note that $\widehat{\boldsymbol{v}}$ in Fig. 3 is the mean of $f(\boldsymbol{v}^{\text{mis}}|\text{data})$ which can be considered a posterior predictive distribution.) Specifically, using $f(v_m^{\text{pred}}|\text{data})$ across m, we produce a map for the posterior predictive SD. Fig. 4, left panel presents such a map for 2006-06-17 (see Supp. Section S1 for the remaining 7 dates). Aside from (a) a slight tendency of higher uncertainty east of 148°E and (b) a cluster of pixels near (150°E, −35.5°N) with highest uncertainty, the AMSR-E uncertainty map shows no recognizable spatial patterns for this date. In fact, the generally patternless nature of the AMSR-E uncertainty map aside from (a) is also evident for 2005-06-13 as well as the remaining 6 dates.

In addition to the SD, 95% credible intervals (CIs) for v_m^{pred} can be presented if desired. Instead, here we present the coverage map of these predictive CIs against v_m^{obs} and use the associated information for model validation. Fig. 4, right panel, presents such a map for 2006-06-17 (see Supp. Section S1 for the remaining 7 dates). According to the 8 maps, there is one recognizable trend, namely, failed coverage typically happens over AMSR-E pixels where either AMSR-E disagrees with ASAR and/or AP, or it disagrees with AP with ASAR missing. By "disagreement," we mean that AMSR-E is high and ASAR and/or AP is low and *vice versa*, relative to the entire MRC. An exception is 2005-06-13, for which failed coverage happens over a region of agreement between AMSR-E and AP (with ASAR missing), roughly where AMSR-E is at least 50 %volume. This can be explained by the fact that, on this date, the three datasets disagree over vast regions of the MRC (see Fig. 1), so that the support of $f(\alpha_1|\text{data})$ is almost entirely negative, and that of $f(\beta_1|\text{data})$ is virtually negative.[6]

For validation of our model in Section 3, we examine the coverage rate summarized in Table 1. Of the 8 dates, three have slight over-coverage and the rest, slight under-coverage. Overall, the near-nominal coverage is one indication that reasonable inference is provided for the MRC. This is true despite that on 2007-06-15 $\widehat{\psi^s}$ values exceed 100 %volume for isolated pixels (see Supp. Section S1). Those anomalies near 147–148.5°E and −36.5°N all correspond to

[5] Very few dates examined had negligible γ_4 and/or γ_5.

[6] All MCMC draws are negative.

Table 1: Coverage of nominal 95% posterior predictive CIs for AMSR-E on dates examined.

Season	Date	# Missing Pixels	Observed Coverage (%)
Summer	2007-01-18	12	96.33
	2008-01-11	22	93.10
	2009-01-22	20	95.55
Winter	2005-06-13	10	93.71
	2006-06-17	1	93.57
	2007-06-15	3	95.47
	2007-06-24	2	93.55
	2007-06-26	0	92.31

the highest AMSR-E values and very high AWAP values, where ASAR is masked. Those anomalies near 147.8°E and −35°N correspond to a missing AMSR-E pixel, where ASAR and AWAP disagree substantially. Nevertheless, out-of-range ψ^s values reveal a limitation of our model: it does not explicitly constrain ψ^s between 0 and 100. Instead, (1) implicitly constrains the mean of ψ^s_{mr} over the mth AMSR-E pixel to be the mean of v_m (which is strictly between 0 and 100). Despite this limitation, v_m^{pred} on 2007-06-15 is well within (20, 75) everywhere (not shown), and the coverage of nominal 95% posterior predictive CIs is very close to 95% (Table 1).

If desired, posterior prediction can be similarly applied to characterize the uncertainty of ASAR and to provide further model validation.

6. Conclusion and future work

As far as we are aware, Chiu and Lehmann (2011) were the first to advocate the use of a hierarchical statistical model for blending multiple soil moisture products. Yet, their work has minimal practical implications in this regard because, admittedly, the influence of ground probe data from 38 sparsely situated stations is hardly noticeable when blended with the areal product, AMSR-E, with wide spatial coverage. In this paper, we have reformulated their modelling framework to blend two areal products, the aforementioned AMSR-E of low spatial resolution with wide coverage, and ASAR, with high resolution but limited coverage on any given day. Compared to Chiu and Lehmann (2011), our results show a heavy influence of ASAR in the blended product, particularly due to the overwhelming volume of ASAR data compared to AMSR-E within an observed region (up to 3600:1). Interestingly, when only 38 randomly selected ASAR pixels were blended with AMSR-E, the resulting blended product resembled that by Chiu and Lehmann (2011) in that they both show features of AMSR-E.

The heavy influence of ASAR swaths on the inference for MRC soil moisture also implies a visible ASAR footprint in the blended product. To investigate if spatially aggregating ASAR would improve the smoothness of the blended product, we modified our model from Section 3 to involve a latent areal ASAR quantity, λ_{mr}, over an AWAP pixel. Specifically, we modified (2) so that $q_{mrk}|\lambda_{mr}, \sigma^2_m \sim \mathrm{N}(\lambda_{mr}, \sigma^2_m)$ and $\lambda_{mr}|\psi_{mr}, \alpha_3, \sigma^2_\varepsilon \sim \mathrm{N}(\alpha_3 \log \psi_{mr}, \sigma^2_\varepsilon)$. Indeed, Fig. 5 shows that this model-based aggregation of ASAR greatly improved the smoothness of the blended product, although at the expense of model complexity due to 312×25 additional model parameters. While the simpler model from Section 3 of this paper produces a blended

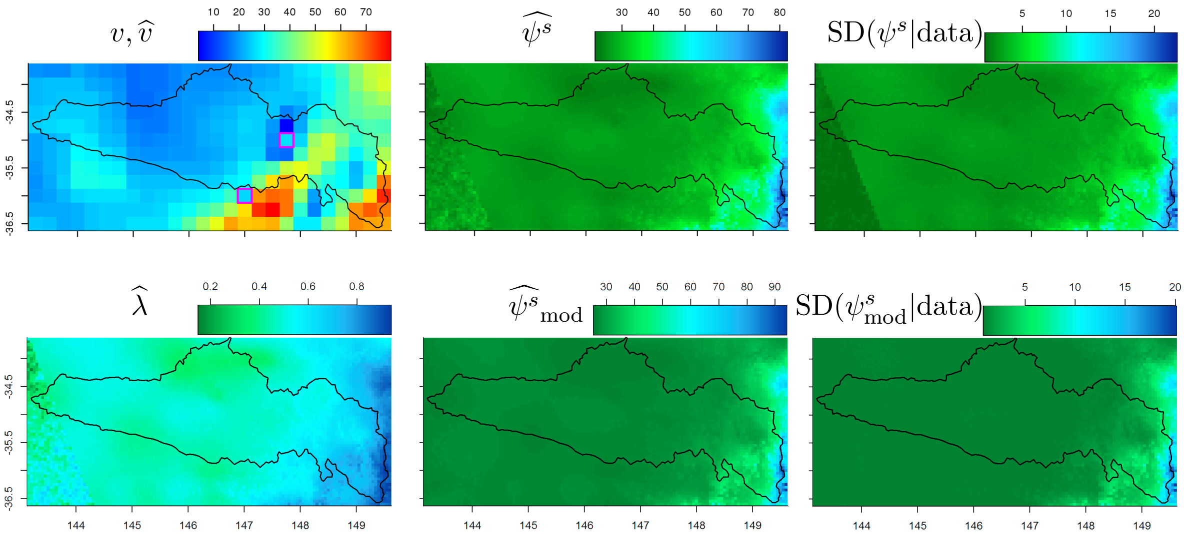

Figure 5: Inference for 2007-06-24 using the model in Section 3 (top row) and that modified to include model-based aggregated ASAR (bottom row). The ASAR footprint is visible from the inference for ψ^s and λ towards the western edge of the rectangular region over the MRC.

product that lacks smoothness, it performs well from a predictive standpoint. Specifically, when judged by the coverage of posterior predictive credible intervals produced for AMSR-E data, the observed credibility was very close to the nominal 95%. Model validation aside, this posterior predictive distribution for AMSR-E allowed us to characterize the uncertainty in satellite data, an objective of high concern to the remote sensing community who relies on satellite measurements of earth-bound physical quantities. The reasonable predictive inference here also reiterates the notion that data perceived as potentially uninformative could be incorporated in a statistical framework to safeguard against biased assimilation, without necessarily sacrificing predictive performance for quantities of interest.

Finally, despite the loss of a popular source of remotely sensed soil moisture data due to the recent failure of AMSR-E, we can anticipate new areal soil moisture products to emerge in the foreseeable future as a result of advancement in remote sensing technology. The modelling framework of this paper then can be adapted to blend areal products, old and new, while taking advantage of relevant covariates as ancillary information. Various directions may be considered when modifying the current framework:

1. Level in the hierarchy at which spatial autocorrelation is modelled. Our current work considers spatial dependence to be fully attributable to the spatial autocorrelation in AWAP data. An alternative approach is to collapse (3) and (4) so that latent SM is modelled as a response of spatial coordinates $\boldsymbol{x}$ alongside AWAP, with spatially autocorrelated residuals. This collapsed model would be easier to implement, although MCMC mixing may worsen unless reparametrizations such as hierarchical centring (Gelfand *et al.* 1995) is employed.

2. Seasonal definition of p_{mr}. As mentioned in Section 3.3, AWAP data may require different transformations for different seasons of the year, depending on which areal products are being modelled.

3. Logit transform of ψ_{mr}. To ensure that ψ^s falls inside (0, 100), a logit transform of latent SM could be considered as an alternative to the logarithmic transform for defining $\tilde{\psi}_{mr}$.

4. Incorporating temporal smoothing. Unifying temporal and spatial smoothing in a single model is substantively more involved than the spatial-only approach of this paper. Chiu (2011) suggests a preliminary spatio-temporal model that is separable and stationary; temporal smoothing is incorporated via sinusoidal basis functions plus an autoregressive-moving-average temporal noise process.

Acknowledgments

The research in this paper was funded by the Water Information Research and Development Alliance between the CSIRO Water for a Health Country Flagship and the Australian Bureau of Meteorology. Model implementation in C++ was funded by eResearch through CSIRO Advanced Scientific Computing. G. Chiu and E. Lehmann thank the Special Issue Editors and Referee, and Drs. S. Bakar and P. Kokic (CSIRO Mathematics, Informatics and Statistics) for their valuable suggestions; Dr. L. Renzullo, Dr. B. Henderson, and fellow team members of the AWRA Model-Data Fusion Project for their ideas and support; and Prof. P. Guttorp (University of Washington) for his invitation to G. Chiu to present the foundational work of this paper at the TIES 2012 Conference.

References

Aubert M, Baghdadi N, Zribi M, Douaoui A, Loumagne C, Baup F, El Hajj M, Garrigues S (2011). "Analysis of TerraSAR-X Data Sensitivity to Bare Soil Moisture, Roughness, Composition and Soil Crust." *Remote Sensing of Environment*, **115**, 1801–1810.

Banerjee S, Carlin BP, Gelfand AE (2004). *Hierarchical Modeling and Analysis for Spatial Data*. Chapman and Hall/CRC, Boca Raton.

Betson RP (1964). "What Is Watershed Runoff?" *Journal of Geophysical Research*, **69**, 1541–1552.

Chiu, GS (2011). "Statistical model-based alternatives to integrate and evaluate soil moisture products." Technical Report #CMIS2011-EP11722. CSIRO Mathematics, Informatics and Statistics.

Chiu GS, Lehmann EA (2011). "Bayesian Hierarchical Modelling: Incorporating Spatial Information in Water Resources Assessment and Accounting." In F Chan, D Marinova, and RS Anderssen (eds.), "MODSIM2011, 19th International Congress on Modelling and Simulation,", ISBN: 978-0-9872143-1-7, URL http://www.mssanz.org.au/modsim2011/.

Cressie N, Wikle CK (2011). *Statistics for Spatio-Temporal Data*. Wiley, Hoboken.

Doubková M, van Dijk AIJM, Sabel D, Wagner W, Blöschl G (2012). "Evaluation of the Predicted Error of the Soil Moisture Retrieval from C-band SAR by Comparison Against Modelled Soil Moisture Estimates over Australia." *Remote Sensing of Environment*, **120**, 188–196.

Eddelbuettel D, François R (2011). "Rcpp: Seamless R and C++ Integration." *Journal of Statistical Software*, **40**, 1–18.

Gelfand AE, Sahu SK, Carlin BP (1995). "Efficient parametrisations for normal linear mixed models." *Biometrika*, **82**, 479–488.

Gelman A, Hill J (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Cambridge University Press, Cambridge.

Gouweleeuw BT, Renzullo LJ, Shoesmith J, Raupach TH (2011). "Data Assimilation and Land Surface Modelling: The Role of Open Source Platforms." In F Chan, D Marinova, and RS Anderssen (eds.), "MODSIM2011, 19th International Congress on Modelling and Simulation," ISBN: 978-0-9872143-1-7, URL http://www.mssanz.org.au/modsim2011/.

Hoff PD (2009). *A First Course in Bayesian Statistical Methods.* Springer, Dordrecht.

Jones DA, Wang W, Fawcett R (2009). "High-Quality Spatial Climate Data-sets for Australia." *Australian Meteorological and Oceanographic Journal*, **58**, 233–248.

Jin H, Henderson B (2011). "Towards a Daily Soil Moisture Product based on Incomplete Time Series Observations of Two Satellites." In F Chan, D Marinova, and RS Anderssen (eds.), "MODSIM2011, 19th International Congress on Modelling and Simulation," ISBN: 978-0-9872143-1-7, URL http://www.mssanz.org.au/modsim2011/.

Kowalczyk EA, Wang YP, Law RM, Davies HL, McGregor JL, Abramowitz G (2006). *The CSIRO Atmosphere Biosphere Land Exchange (CABLE) Model for Use in Climate Models and as an Offline Model.* Paper 013. CSIRO Marine and Atmospheric Research.

Lindgren F, Rue H, 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 B*, **73**, 423–498.

Lord CG, Taylor CA (2009). "Biased Assimilation: Effects of Assumptions and Expectations on the Interpretation of New Evidence." *Social and Personality Psychology Compass*, **3**, 827–841.

Mladenova I, Lakshmi V, Walker JP, Panciera R, Wagner W, Doubkova M (2010). "Validation of the ASAR Global Monitoring Mode Soil Moisture Product Using the NAFE'05 Data Set." *IEEE Transactions on Geoscience and Remote Sensing*, **48**, 2498–2508.

Owe M, de Jeu RAM, Holmes TRH (2008). "Multi-Sensor Historical Climatology of Satellite-Derived Global Land Surface Moisture." *Journal of Geophysical Research*, **113**, F01002.

Pathe C, Wagner W, Sabel D, Doubkova M, Basara JB (2009). "Using ENVISAT ASAR Global Mode Data for Surface Soil Moisture Retrieval over Oklahoma, USA." *IEEE Transactions on Geoscience and Remote Sensing*, **47**, 468–480.

Renzullo LJ, van Dijk AIJM, Gouweleeuw BT (2012). "On the modelling of microwave brightness temperatures: Implications for satellite data assimilation into the Australian Water Resources Assessment system." In "Water Information Research and Development Alliance: Science Symposium Proceedings, Melbourne, Australia, 1–5 August 2011," ISBN 978-0-643-10825-7, URL http://www.csiro.au/Portals/Publications/WIRADA_WfHC_Pub.aspx.

van Dijk AIJM, Renzullo LJ (2011). "Water Resource Monitoring Systems and the Role of Satellite Observations." *Hydrology and Earth System Sciences*, **15**, 39–55.

van Dijk A, Warren G (2010). *The Australian Water Resources Assessment System Report 4.0.5.: Evaluation Against Observations.* Water for a Healthy Country Flagship. CSIRO.

Wigneron JP, Calvet JC, Pellarin T, van de Griend AA, Berger M, Ferrazzoli P (2003). "Retrieving Near-Surface Soil Moisture from Microwave Radiometric Observations: Current Status and Future Plans." *Remote Sensing of Environment*, **85**, 489–506.

Appendices

A1. MCMC algorithm and implementation

Below is our overall MH-within-Gibbs sampling algorithm for approximating (7). The index t denotes the tth MCMC scan.

1. Set t=0. Specify reasonable starting values $\boldsymbol{\Omega}^{(0)}, \boldsymbol{\phi}^{(0)}$, and $\boldsymbol{\psi}^{(0)}$ (see Supp. Section S2).
2. Increment t. If the Markov chain has not converged, then do Steps 3–8.
3. Sample $\boldsymbol{v}^{\text{mis}(t)}$ according to (8) with relevant parameters from the preceding t.
4. Do Step 4a in parallel over m=1, ..., M.
   (a) (Note that this step affects $c_{12,m,r+1}$.) For r=1, ..., 25, perform the following:
      i. propose ψ^*_{mr} by sampling from (20) with $z = \psi^{(t-1)}_{mr}$ where the superscript $(t-1)$ corresponds to the preceding t;
      ii. compute R_ν by exponentiating (22) and substituting in it relevant parameters from (I) the preceding t, (II) Step 3, and (III) if r>1, then also Step 4a from the preceding r;
      iii. sample $U \sim \text{Unif}[0, 1]$ and set

$$\psi^{(t)}_{mr} = \begin{cases} \psi^*_{mr} \text{ (i.e. accept proposed)} & \text{if } R_\nu > U \\ \psi^{(t-1)}_{mr} \text{ (i.e. reject proposed)} & \text{otherwise} \end{cases}.$$

5. In parallel with Step 4, do as follows.
   (a) (Note that this step affects $c_{01,i+1}$.) For i=1, ..., 25M, sample $\phi^{(t)}_i$ according to (10) with relevant parameters from (i) the preceding t and (ii) if i>1, then also Step 5a from the preceding i.
   (b) Centre $\phi^{(t)}_1, \ldots, \phi^{(t)}_{25M}$ by applying the reassignment $\phi^{(t)}_i \leftarrow \phi^{(t)}_i - \overline{\phi}^{(t)}$.
6. Do the following in parallel:
   (a) Sample $\alpha^{(t)}_3$ according to (11) with relevant parameters from (i) the preceding t and (ii) Steps 3–4.
   (b) Sample $[\alpha^{(t)}_0, \alpha^{(t)}_1]'$ according to (12) with relevant parameters from (i) the preceding t and (ii) Steps 3–4.
   (c) Sample $\boldsymbol{\beta}^{(t)}$ according to (13) with relevant parameters from (i) the preceding t and (ii) Step 4.
   (d) Sample $\boldsymbol{\gamma}^{(t)}$ according to (14) with relevant parameters from (i) the preceding t and (ii) Step 5b.
   (e) Sample $\tau^{2(t)}$ according to (15) with relevant parameters from Step 5b.
7. Do the following in parallel:
   (a) For each m, sample $\sigma^{2(t)}_m$ according to (16) with relevant parameters from Steps 3, 4, and 6b.
   (b) Sample $\sigma^{2(t)}_\delta$ according to (17) with relevant parameters from Steps 3, 4, and 6a.
   (c) Sample $\sigma^{2(t)}_\eta$ according to (18) with relevant parameters from Steps 4 and 6c.
   (d) Sample $\sigma^{2(t)}_\zeta$ according to (19) with relevant parameters from Steps 5b and 6d.
8. Return to Step 2.

MCMC implementation and data analyses were performed in R 2.14 and above (serial and parallel) and C++ (serial). For sparse matrix algebra and parallelization in R, we used the packages **Matrix** (http://r-forge.r-project.org/projects/matrix) and **snow** (http://cran.r-project.org/web/packages/snow/index.html), respectively. For our data, we identified that parallelization in R within Step 4 led to major reduction in computational time, while parallelization across and within the other steps often increased computational overhead. We subsequently incorporated C++ into Steps 4 and 5 of the original R implementation using the package **Rcpp** (Eddelbuettel and François 2011). The directed use of C++ gave the highest performance code, with an approximate speedup of 20 times the serial version in R. Parallelization of the C++ code was deemed unnecessary by the compiler's automatic parallelizer.

A2. AMSR-E posterior predictive distribution

Let $\boldsymbol{\theta}=\{\boldsymbol{v}^{\text{mis}}, \boldsymbol{\psi}, \boldsymbol{\phi}, \boldsymbol{\Omega}\}$. Then,

$$f\left(\boldsymbol{v}^{\text{pred}}\middle|\,\text{data}\right) = \int_{\boldsymbol{\theta}} f\left(\boldsymbol{v}^{\text{pred}}\middle|\,\boldsymbol{\theta},\text{data}\right) f\left(\boldsymbol{\theta}|\,\text{data}\right)\, d\boldsymbol{\theta} = \int_{\boldsymbol{\theta}} f\left(\boldsymbol{v}^{\text{pred}}\middle|\,\boldsymbol{\theta}\right) f\left(\boldsymbol{\theta}|\,\text{data}\right)\, d\boldsymbol{\theta} \quad (23)$$

where $f\left(\boldsymbol{\theta}|\,\text{data}\right)$ is equivalent to (7). We approximate (23) by making use of the MCMC approximation of (7), as follows. Given the tth draw $\boldsymbol{\theta}^{(t)}$ from $f(\boldsymbol{\theta}|\,\text{data})$, sample $\boldsymbol{v}^{\text{pred}(t)}$ from $f(\boldsymbol{v}|\,\boldsymbol{\theta}^{(t)})$ according to (1), i.e. sample

$$\boldsymbol{v}^{\text{pred}(t)} \sim \text{MVN}\left(\alpha_0^{(t)} + \alpha_1^{(t)}\overline{\boldsymbol{\psi}}^{(t)}, (\sigma_\delta^{(t)})^2\mathbb{I}_M\right).$$

The set $\{\boldsymbol{v}^{\text{pred}(t)} : t=1, 2, \ldots\}$ constitutes an approximation to (23). (See, e.g. Hoff 2009, for details of MCMC approximations for posterior predictive distributions).

Affiliation:

Grace S. Chiu (corresponding author)
CSIRO Mathematics, Informatics and Statistics
GPO Box 664, Canberra, ACT 2601, Australia
E-mail: grace.chiu@csiro.au
URL: http://www.csiro.au/people/Grace.Chiu.html

Eric A. Lehmann
CSIRO Mathematics, Informatics and Statistics
Private Bag 5, Wembley, WA 6913, Australia
E-mail: eric.lehmann@csiro.au

Joshua C. Bowden
CSIRO Information Management and Technology
306 Carmody Road, St. Lucia, QLD 4067, Australia
E-mail: josh.bowden@csiro.au

Journal of Environmental Statistics http://www.jenvstat.org
Volume 4, Issue 9 Submitted: 2012-03-29
February 2013 Accepted: 2012-10-10