svc: Spatially Varying Coefficient Models in R

Justice Akuoko-Frimpong, Edward Shao, Jonathan Ta

1 Introduction

Heterogeneity in spatial processes poses challenges for traditional regression models. For example, consider how the relationship between PM2.5 concentrations and factors like traffic density or industrial emissions vary across an urban-rural gradient. Global regression models, assuming constant coefficients, can misrepresent these spatially localized relationships, leading to incorrect conclusions and suboptimal policies.

We propose a Bayesian approach for spatially varying coefficient (SVC) models to address these issues by allowing coefficients to change smoothly over space. Our framework enhances efficiency through three key innovations: a low-rank Gaussian process approximation, a robust adaptive Metropolis (RAM) for optimal random walk sampling, and optimized linear algebra using pre-computed distance matrices and Cholesky decompositions. These improvements make SVC models feasible for large-scale environmental datasets.

2 Methods

Let $\mathcal{D} \subset \mathrm{R}^2$ denote a spatial domain, $s_i \in \mathcal{D}$ for each $i = 1, \dots, n$ denote a spatial location with collected data, and $\boldsymbol{s} = (s_1, \dots, s_n)^\top$ be the vector of all such locations. Then $Y(\boldsymbol{s})$ are univariate dependent variables and $\boldsymbol{X}(\boldsymbol{s}) = (X_1(\boldsymbol{s}), \dots, X_p(\boldsymbol{s}))^\top$ are $p \times 1$ vectors of covariates. A linear SVC regression model assumes $Y(\boldsymbol{s})$ are dependent on $\boldsymbol{X}(\boldsymbol{s})$ as follows:

$$Y(\boldsymbol{s}) = \sum_{j=1}^{p} X_r(\boldsymbol{s}) w_r(\boldsymbol{s}) + \epsilon(\boldsymbol{s}),$$

where $w_r(\boldsymbol{s})$ are the SVC's corresponding to $X_r(\boldsymbol{s})$ and $\epsilon(\boldsymbol{s})$ are independently and identically distributed multivariate normal measurement errors, i.e. $\epsilon(\boldsymbol{s}) \sim \text{MVN}(0, \tau^2 I_n)$. Note that spatially-varying intercept can be included by defining $X_1(\boldsymbol{s})$ as $\mathbf{1}_n$, a vector of length n with every element equal to 1.

For each $w_r(\boldsymbol{s})$, we assign a Gaussian process (GP) prior with squared exponential covariance, i.e. $w_r \sim \text{GP}(0, C(\boldsymbol{\theta}_r))$ where $C(\boldsymbol{\theta}_r) = [C(s_i, s_{i'}; \boldsymbol{\theta}_r)]_{i,i'=1}^n$. We refer to $C(\cdot)$ as the covariance function, because the covariance of $w_r(s_i)$ and $w_r(s_{i'})$ at locations s_i and $s_{i'}$ is $\sigma_r^2 C(\boldsymbol{\theta}_r)_{i,i'}$. In particular, we use the squared exponential covariance function where $\boldsymbol{\theta}_r = (\sigma_r^2, \phi_r)$, $C(\boldsymbol{\theta}_r) = \sigma_r^2 K(\phi_r)$, and $K(\phi_r) = [K(s_i, s_{i'}; \phi_r)]_{i,i'=1}^n = [\exp\{\phi_r^{-1} \| s_i - s_{i'} \|^2\}]_{i,i'=1}^n$, due to its useful properties such as infinite smoothness and gradual decrease in covariance with distance. The parameter σ_r^2 is the spatial variance and ϕ_r is the spatial decay, which indicates how quickly correlation decreases with the squared distance.

We assign inverse-gamma conjugate priors for the variance parameters τ^2 and each σ_r^2. In addition, each ϕ_r is assigned a uniform prior. In summary, $\tau^2 \sim \text{Inv.Gamma}(\alpha_\tau, \beta_\tau)$, $\sigma_r^2 \sim \text{Inv.Gamma}(\alpha_r, \beta_r)$, and $\phi_r \sim \text{Uniform}(l_r, u_r)$.

2.1 MCMC Algorithm

The primary function in the `svc` package, `svclm`, samples from the joint posterior distribution of $w_r(\boldsymbol{s}), \phi_r, \sigma_r^2$, and τ^2, for $r = 1, \ldots, p$ using a Gibbs sampler. We will explain how each parameter is sampled at iteration $t + 1$.

The parameter $\phi_r^{(t+1)}$ is updated via a random walk Metropolis algorithm. Because

$\phi_r^{(t)} \in (l_r, u_r)$, each proposal is calculated as

$$\phi_r' = f^{-1}(f(\phi_r^{(t)}) + U) = l_r + \frac{u_r - l_r}{1 + \exp\left\{\log\left(\frac{u_r - l_r}{\phi_r^{(t)} - l_r} - 1\right) + U\right\}},$$

where U is sampled from a normal distribution with mean 0. By default, the proposal standard deviation starts at 1 and adapts via the RAM algorithm to achieve a target acceptance rate as described by Vihola (2012).

The sample the SVC's, let $\widetilde{Y}(\boldsymbol{s}) = Y(\boldsymbol{s}) - \sum_{j \neq r} X_j(\boldsymbol{s}) w_j^{(t)}(\boldsymbol{s})$. Then

$$\frac{\widetilde{Y}}{X_r}(\boldsymbol{s})|w_r^{(t)}(\boldsymbol{s}), \tau^2 \sim \text{MVN}\left(w_r^{(t)}(\boldsymbol{s}), \tau^{2(t)}\text{diag}\left(\frac{1}{X_r^2(\boldsymbol{s})}\right)\right) := \text{MVN}\left(\mu, \Sigma\right).$$

Then assuming the prior $w_r(\boldsymbol{s}) \sim \text{MVN}\left(0, C(\boldsymbol{\theta}_r^{(t)})\right) := \text{MVN}\left(0, \Sigma_0\right)$, we sample $w_r^{(t+1)}(\boldsymbol{s})$ from MVN(μ_1, Σ_1), where $\Sigma_1 = \left(\Sigma_0^{-1} + \Sigma^{-1}\right)$ and $\mu_1 = \Sigma_1 \Sigma^{-1} \frac{\widetilde{Y}}{X_r}(s)$.

We sample $\sigma_r^{2(t+1)}$ from Inv.Gamma$\left(\alpha_r + \frac{n}{2}, \beta_r + w_r^{(t)}(\boldsymbol{s})^\top K(\phi_r^{(t)})^{-1} w_r^{(t)}(\boldsymbol{s})\right)$ and $\tau^{2(t+1)}$ from Inv.Gamma$\left(\alpha_t + \frac{n}{2}, \beta_t + \frac{1}{2}\left(Y(\boldsymbol{s}) - \sum_{j=0}^{p} X_r(\boldsymbol{s}) w_r^{(t)}(\boldsymbol{s})\right)\right)$.

2.1.1 Low-Rank Gaussian Process

We can perform the above algorithm using the complete data for all n locations to estimate the full-rank Gaussian process model. The most flop-expensive operation is inverting $C(\boldsymbol{\theta}_r^{(t)})$, an $(n \times n)$ matrix, for $r = 1, \ldots, p$, resulting in an overall cost of $O(pn^3)$ per iteration. We can improve the efficiency of this algorithm with a low-rank Guassian process as described by Banerjee et al. (2008).

Consider a set of knots $\boldsymbol{s}^* = (s_1^*, \ldots, s_m^*)^\top$ which are a subset of locations $\boldsymbol{s}$. With the low-rank Guassian process, we sample τ^2, ϕ_r, σ_r^2, and $\boldsymbol{w}_r^* = [w(\boldsymbol{s}_i^*)]_{i=1}^m$ which follows a multivariate Gaussian distribution with covariance matrix $C^*(\boldsymbol{\theta}_r) = \sigma_r^2 K^*(\phi_r) = [\sigma_r^2 \exp\{\phi_r^{-1}\|s_i^* - s_i^*\|^2\}]_{i,i'=1}^m$, for each $r = 1, \ldots, p$.

Then for the original set of locations $\boldsymbol{s}$, we predict the spatial interpolants $\widetilde{w}_r(\boldsymbol{s}) =$

$\mathrm{E}[w_r(\boldsymbol{s})|\boldsymbol{w}_r^*] = \boldsymbol{c}(\boldsymbol{s};\boldsymbol{\theta}_r)C^{*-1}(\boldsymbol{\theta}_r)\boldsymbol{w}_r^*$, where $\boldsymbol{c}(\boldsymbol{s};\boldsymbol{\theta}_r)$ is an $n \times m$ matrix with i, j element $c_{i,j} = \sigma_r^2 K(s_i, s_j^*; \phi_r) = \sigma_r^2 \exp\{\phi_r^{-1}\|s_i - s_j^*\|^2\}$.

Therefore, using the low-rank Gaussian process, we can replace $w_r(\boldsymbol{s})$ in the original model with $\widetilde{w}_r(\boldsymbol{s})$ for each $r = 1, \ldots, p$, but with an overall cost per iteration of $O(pm^3)$, where $m < n$.

3 Using svc

To fit a linear SVC regression model using `svc`, we use the `svclm` function. In most cases, we recommend providing the following arguments to `svclm`: `Y`, `X`, `coords`, `Y_knots`, `X_knots`, `knots`, `phi_lower`, `phi_upper`, and `mcmc`. `Y` is a vector of length n containing the response variable, `X` is the design matrix of size $n \times p$, and `coords` is a matrix of size $n \times 2$ containing the spatial coordinates of each location. The arguments `Y_knots`, `X_knots`, and `knots` provide the data for the knots that will be used in the low-rank model. `Y_knots` is a vector of length m containing the response variable at the knots, `X_knots` is a matrix of size $m \times p$ containing the covariates at the knots, and `knots` is a matrix of size $m \times 2$ containing the spatial coordinates of each knot. The arguments `phi_lower` and `phi_upper` are each a vector of length p containing the lower and upper bounds for the uniform distribution priors assigned to spatial decay parameters ϕ_r, respectively. The argument `mcmc` is an integer specifying the number of MCMC iterations to run.

The arguments above are sufficient for most cases, but `svclm` also has arguments for specifying starting values for τ^2, σ_r^2, ϕ_r, and $\boldsymbol{w}_r^*$ where $r = 1, \ldots, p$. Users can also set the hyperparameters for the inverse gamma priors on τ^2 and σ_r^2. Lastly, users can set the starting proposal standard deviations for each ϕ_r and the target acceptance rate for the RAM.

By default, the starting values for τ^2, σ_r^2, ϕ_r, and $\boldsymbol{w}_r^*$ are set to 1, 1, the midpoint between the corresponding upper and lower bounds, and 0 for all locations, respectively. The default priors for τ^2 and each σ_r^2 are all Inv.Gamma$(0.001, 0.001)$ to achieve uninformative priors. The starting proposal standard deviations for each ϕ_r are set to 1 by default, and the target acceptance rate for the RAM algorithm is set to 0.234 which is asymptotically optimal under most conditions (Gelman et al. 1997).

In all cases, `svclm` returns a list containing the following matrices: `w_samples`, an ($\texttt{mcmc} \times n \times p$) array of the posterior samples for the SVC's; `phi_samples`, an ($\texttt{mcmc} \times p$) matrix of the posterior samples for the spatial decay parameters; `phi_acceptance`, an ($\texttt{mcmc} \times p$) matrix which is 1 if the proposal for ϕ_r was accepted and 0 otherwise; `sigmasq_samples`, an ($\texttt{mcmc} \times p$) matrix of the posterior samples for the spatial variances; and `tausq_samples`, a ($\texttt{mcmc} \times 1$) matrix of the posterior samples for the nugget parameter.

Although knot data can be chosen manually, the `svc` package also contains a function `simpleknots` to automatically select knots for the low-rank model. The function takes arguments `Y`, `X`, `coords`, and `k`, where `k` is an integer and `Y`, `X`, and `coords` match what would be provided to the `svclm` function. The function will return a list containing the objects `Y_knots`, `X_knots`, and `knots` which can then be passed directly to the `svclm` function. The knots are selected by sorting each dimension of coordinates and then selecting every k-th coordinate in each dimension.

4 Simulation

We compared the performance of our package with **spBayes** using bias, RMSE, and computational time (in seconds), averaged over 1000 simulation replications. In each simulation, we generated data for 441 locations using the same parameters and obtained knot data using the `simpleknots` function with `k` set to 2. Then we used both packages to run an MCMC for 3000 iterations and predicted the coefficients at all locations using the last 1000 iterations of each MCMC chain.

Each dataset was generated with $p = 3$, $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 1$, $\phi_1 = \phi_2 = \phi_3 = 2$, and $\tau^2 = 0.0001$. The covariates were generated as $X_1 = \mathbf{1}_n$, $X_2 \sim \mathcal{N}(0,1)$, and $X_3 \sim \mathcal{N}(0,1)$. True values for the means of w_0, w_1, and w_2 were set to 0, 10, and -5 respectively. The response variable was generated as $Y(\boldsymbol{s}) = w_1(\boldsymbol{s}) + w_2(\boldsymbol{s})X_2 + w_3(\boldsymbol{s})X_3 + \epsilon(\boldsymbol{s})$.

Table 1 summarizes the results of the simulation. The **svc** package had better bias and RMSE for w_1 but worse bias and RMSE for w_2 and w_3. Our package also ran more than 3 times faster on average for the same number of iterations.

Table 1: Comparison of methods based on bias, RMSE, and computation time (seconds).

Method	w_1		w_1		w_2		**Time (s)**
	Bias	RMSE	Bias	RMSE	Bias	RMSE	
svc (low rank)	0.00	0.09	1.16	1.48	-1.30	2.08	7.17
spBayes (low rank)	-0.01	0.21	-0.01	0.25	0.01	0.22	23.26

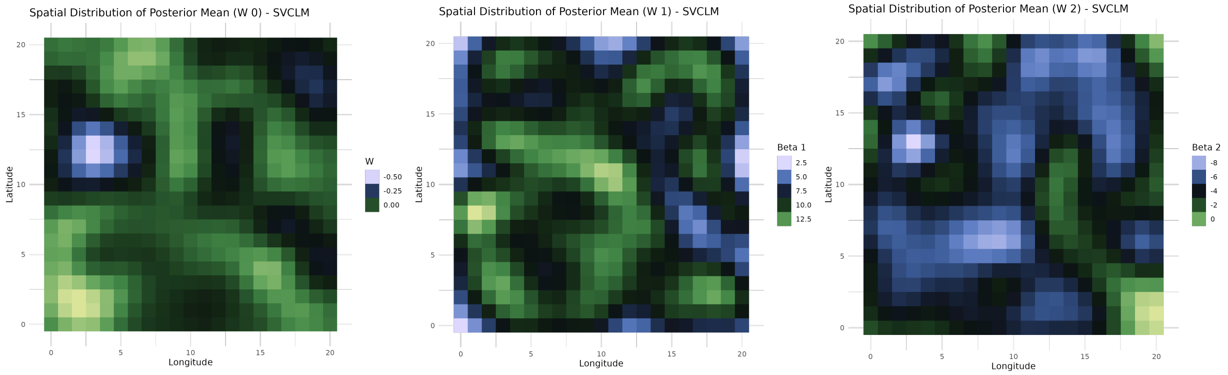

Figure 1: Posterior Mappings for Coefficients - SVCLM low rank

5 Data Analysis

The real world data used in this section is a gridded satellite data with latitude, longitude, temperature, Normalized Difference Vegetation Index (NDVI); which measures greenness, and emissivity which represents a surface's efficiency in emitting thermal radiation (Hulley & Hook 2008, Hulley et al. 2008, Hulley & Hook 2009, Hulley et al. 2009, Hulley & Hook 2011, Hulley et al. 2012, NASA JPL 2014, Hulley et al. 2015). The data had 111,556 observations and 6 variables namely, latitude, longitude, emissivity, NDVI and land surface tempareture (LST). Variables in the dataset were means computed from all clear-sky pixels of the Terra Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) scenes acquired from 2000 through 2008. We used LST as a univariate outcome and used emissivity and NDVI as covariates. Figure 2 below reveals the spatial distributions of the variables in the dataset.

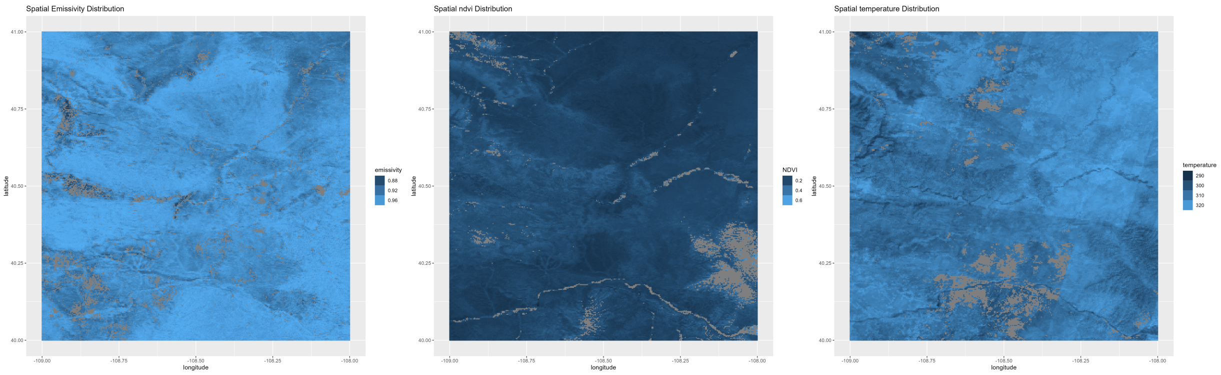

Figure 2: Spatial distributions of emissivity, NDVI, and land surface temperature in the study area.

Emissivity ranged from 0.88 to 0.96, NDVI ranged from 0.2 to 0.6 and LST, from 290 to 320. Cooler regions often align with greener (higher NDVI) areas — vegetation can reduce surface temperature through evapotranspiration.

5.1 Model

We modeled (LST) using an SVC framework:

$$\text{Temp}(\boldsymbol{s}) = w_1(\boldsymbol{s}) + w_2(\boldsymbol{s}) \cdot \text{NDVI} + w_3(\boldsymbol{s}) \cdot \text{Emissivity} + \epsilon(\boldsymbol{s}),$$

with priors as specified in Section 2.

We specified weakly informative priors: $\tau^2 \sim \text{Inv-Gamma}(0.001, 0.001)$, $\sigma_r^2 \sim \text{Inv-Gamma}(0.001, 0.001)$, and $\phi_r \sim \text{Uniform}(0.001, 500)$, for each $r = 1, \ldots 3$.

The computational implementation featured low-rank approximation. We selected 1 knot per 100 grid cells ($k = 10$) using the `simpleknots()` function. Removed knots with missing data.

5.2 Results

Below is a plot of the posterior means of the coefficients at each location. The means were computed after burn-in.

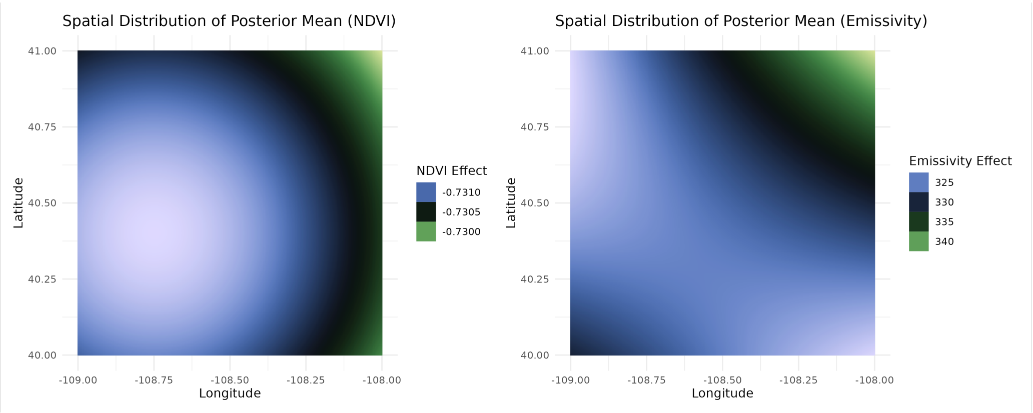

Figure 3: Spatially varying coefficient surfaces for (A) NDVI and (B) Emissivity effects on land surface temperature

The posterior mean NDVI coefficients show strong negative effects (cooling) across the entire study region, with values ranging from -0.7310 to -0.7300. Each unit increase in

NDVI corresponds to about $0.73°C$ decrease in land surface temperature across space. The emissivity coefficients ranged from 325 to 340. These values revealed strong positive associations which implies that across space, a unit increase in emissivity increases temperature.

We generated predicted temperature values across the entire spatial domain using the posterior mean estimates of the SVC's. These predictions were computed for all locations with complete predictor data, regardless of whether the actual temperature was observed. The predicted values were obtained by multiplying the mean coefficient estimates with the corresponding covariate values at each location. We then visualized both the observed and predicted temperature surfaces side by side to assess model performance and spatial prediction accuracy.

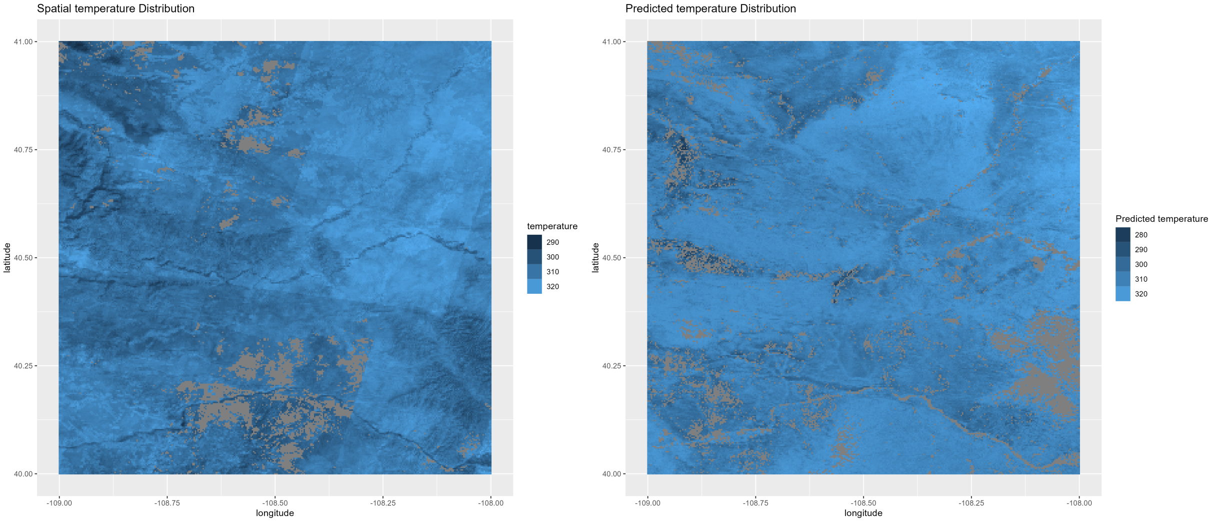

Figure 4: Side by side plot of temperature and their predicted values across space. Gray indicates missing values.

Note that the gray spaces in the plot are the locations with missing data. We can predict LST at locations with missing LST but cannot predict at locations with missing covariates; NDVI and emissivity.The figure comparison reveals the model's ability to capture spatial patterns in LST, even in regions where observations were missing but predictor data were available. The range of values between the actual and predicted were close.

6 Conclusion

The **svc** package provides a flexible and efficient framework for fitting SVC models that capture heterogeneous spatial processes. By leveraging low-rank Gaussian process approximations, RAM adaptation, and various sampling optimizations, we have made SVC models more computationally feasible for large-scale spatial datasets.

Compared to **spBayes**, our method performs more efficiently and has lower bias and RMSE for the spatial random intercept, but has higher bias and RMSE for all other SVC's. However, **svc** did run much faster than **spBayes** for the same number of iterations.

One limitation of our method is that it requires complete data at knot locations, whereas **spBayes** does not require any data at selected knot locations.

Further improvements to the package could include allowing for different covariance functions and implementing parallelization for sampling ϕ_r and σ_r^2 for $r = 1, \ldots, p$ since the sampling can be done independently for each r.

References

Banerjee, S., Gelfand, A. E., Finley, A. O. & Sang, H. (2008), 'Gaussian predictive process models for large spatial data sets', **70**(4), 825–848.

URL: *https://academic.oup.com/jrsssb/article/70/4/825/7109503* 3

Gelman, A., Gilks, W. R. & Roberts, G. O. (1997), 'Weak convergence and optimal scaling of random walk metropolis algorithms', **7**(1).

URL: *https://projecteuclid.org/journals/annals-of-applied-probability/volume-7/issue-1/Weak-convergence-and-optimal-scaling-of-random-walk-Metropolis-algorithms/10.1214/aoap/1034625254.full* 5

Hulley, G. C. & Hook, S. J. (2008), 'A new methodology for cloud detection and classification with ASTER data', **35**(16).

URL: *https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2008GL034644* 7

Hulley, G. C. & Hook, S. J. (2009), 'The north american ASTER land surface emissivity database (NAALSED) version 2.0', **113**(9), 1967–1975.

URL: *https://linkinghub.elsevier.com/retrieve/pii/S0034425709001631* 7

Hulley, G. C. & Hook, S. J. (2011), 'Generating consistent land surface temperature and emissivity products between ASTER and MODIS data for earth science research', **49**(4), 1304–1315.

URL: *http://ieeexplore.ieee.org/document/5582278/* 7

Hulley, G. C., Hook, S. J., Abbott, E., Malakar, N., Islam, T. & Abrams, M. (2015), 'The <span style="font-variant:small-caps;">ASTER</span> global emissivity dataset (<span style="font-variant:small-caps;">ASTER GED</span>): Mapping earth's emissivity at 100 meter spatial scale', **42**(19), 7966–7976.

URL: *https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2015GL065564* 7

Hulley, G. C., Hook, S. J. & Baldridge, A. M. (2008), 'ASTER land surface emissivity database of california and nevada', **35**(13), 2008GL034507.

URL: *https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2008GL034507* 7

Hulley, G. C., Hook, S. J. & Baldridge, A. M. (2009), 'Validation of the north american ASTER land surface emissivity database (NAALSED) version 2.0 using pseudo-invariant sand dune sites', **113**(10), 2224–2233.

URL: *https://linkinghub.elsevier.com/retrieve/pii/S0034425709001849* 7

Hulley, G. C., Hughes, C. G. & Hook, S. J. (2012), 'Quantifying uncertainties in land surface temperature and emissivity retrievals from ASTER and MODIS thermal infrared data', **117**, 2012JD018506.

URL: *https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2012JD018506* 7

NASA JPL (2014), 'ASTER global emissivity dataset, 100-meter, HDF5'.

URL: *https://lpdaac.usgs.gov/products/ag100v003/* 7

Vihola, M. (2012), 'Robust adaptive metropolis algorithm with coerced acceptance rate', **22**(5), 997–1008.

URL: *https://doi.org/10.1007/s11222-011-9269-5* 3