

Supplementary Materials for *Locally Adaptive Smoothing with Markov Random Fields and Shrinkage Priors*

James R. Faulkner and Vladimir N. Minin

1 Approximation to the Horseshoe Density

There is no exact closed-form expression available for the horseshoe density function. We present an approximation to the horseshoe density that can be used without the need for explicit specification of the nuisance local scale parameters. Following Carvalho et al. (2010), the marginal distribution of u given global scale parameter γ is found by integrating over possible values of the local scale parameter τ , where $u|\tau \sim N(0, \delta\tau^2)$ and $\tau|\gamma \sim C^+(0, \gamma)$. Here δ is a constant representing a scale factor for the distance between adjacent points when this distribution is used for the increments of a k th-order smoothing model. This leads to

$$\begin{aligned} p(u|\delta, \gamma) &= \int_0^\infty p(u|\delta, \tau, \gamma)p(\tau|\gamma)d\tau \\ &= \int_0^\infty \frac{1}{\sqrt{2\pi\delta\tau^2}} \exp\left(-\frac{u^2}{2\delta\tau^2}\right) \frac{2\gamma}{\pi(\tau^2 + \gamma^2)} d\tau \end{aligned}$$

We let $B = 2\gamma/(\sqrt{2\pi^3}\delta)$ and introduce the substitution $\omega = \tau^2$, which gives $d\tau = -1/(2\omega^{3/2})$, resulting in

$$\begin{aligned} p(u|\delta, \gamma) &= B \int_0^\infty \frac{\omega^{1/2}}{2\omega^{3/2}} \exp\left(-\frac{u^2\omega}{2\delta}\right) \frac{1}{\omega^{-1} + \gamma^2} d\omega \\ &= \frac{B}{2} \int_0^\infty \frac{1}{\omega} \exp\left(-\frac{u^2\omega}{2\delta}\right) \frac{\omega}{1 + \omega\gamma^2} d\omega \\ &= \frac{B}{2} \int_0^\infty \frac{1}{1 + \omega\gamma^2} \exp\left(-\frac{u^2\omega}{2\delta}\right) d\omega. \end{aligned}$$

Now we introduce the substitution $z = 1 + \omega\gamma^2$, which gives $d\omega = \gamma^{-2}dz$, and results in

$$\begin{aligned}
p(u|\delta, \gamma) &= \frac{B}{2} \int_1^\infty \frac{1}{z\gamma^2} \exp\left\{-\left(\frac{u^2\omega}{2\delta}\right)\frac{z}{\gamma^2} + \left(\frac{u^2}{2\delta}\right)\frac{1}{\gamma^2}\right\} dz \\
&= \frac{B}{2\gamma^2} \exp\left(\frac{u^2}{2\delta\gamma^2}\right) \int_1^\infty \frac{1}{z} \exp\left(-\frac{u^2z}{2\delta\gamma^2}\right) dz \\
&= \left(\frac{1}{2\pi^3\delta\gamma^2}\right)^{1/2} \exp\left(\frac{u^2}{2\delta\gamma^2}\right) E_1\left(\frac{u^2}{2\delta\gamma^2}\right) dz,
\end{aligned}$$

where E_1 is the exponential integral function. Note that $\lim_{x \rightarrow 0^+} E_1(x) = \infty$, but for $x > 0$, the function $E_1(x)$ is bounded as follows:

$$\frac{1}{2}e^{-x} \ln\left(1 + \frac{2}{x}\right) < E_1(x) < e^{-x} \ln\left(1 + \frac{1}{x}\right).$$

Then for $u \in \{\mathbb{R} : u \neq 0\}$ we have

$$\frac{1}{2} \exp\left(\frac{-u^2}{2\delta\gamma^2}\right) \ln\left(1 + \frac{4\delta\gamma^2}{u^2}\right) < E_1\left(\frac{u^2}{2\delta\gamma^2}\right) < \exp\left(\frac{-u^2}{2\delta\gamma^2}\right) \ln\left(1 + \frac{2\delta\gamma^2}{u^2}\right).$$

It follows that the target density is bounded by

$$\frac{1}{2} \left(\frac{1}{2\pi^3\delta\gamma^2}\right)^{1/2} \ln\left(1 + \frac{4\delta\gamma^2}{u^2}\right) < p(u|\gamma) < \left(\frac{1}{2\pi^3\delta\gamma^2}\right)^{1/2} \ln\left(1 + \frac{2\delta\gamma^2}{u^2}\right). \quad (1)$$

Let the left bound in equation (1) be denoted $B_1(u)$ and the right bound $B_2(u)$. Note that as $u \rightarrow 0$, each of $B_1(u)$, $p(u|\gamma)$ and $B_2(u)$ approach ∞ . It can be shown that $\int_{-\infty}^\infty B_1(u)du = \sqrt{2/\pi}$ and $\int_{-\infty}^\infty B_2(u)du = 2/\sqrt{\pi}$. Since $\sqrt{2/\pi} < 1 < 2/\sqrt{\pi}$, these bounds can be used to find an approximate expression for $p(u|\gamma)$ that integrates to 1 and still satisfies equation (1). We set

$$\tilde{p}(u|\gamma) = wB_1(u) + (1-w)B_2(u) \quad (2)$$

with constraints $0 < w < 1$ and $\int_{-\infty}^\infty wB_1(u) + (1-w)B_2(u)du = 1$. Using the values for the integrated bounds and solving gives $w = (\sqrt{\pi} - 2)/(\sqrt{2} - 2)$. Substituting this value for w into equation (2) and simplifying gives the following closed-form approximation to the horseshoe density function:

$$\tilde{p}(u|\gamma) = \left(\frac{1}{2\pi^3\delta\gamma^2}\right)^{1/2} \left[\frac{\sqrt{\pi} - 2}{2\sqrt{2} - 4} \ln\left(1 + \frac{4\delta\gamma^2}{u^2}\right) + \frac{\sqrt{2} - \sqrt{\pi}}{\sqrt{2} - 2} \ln\left(1 + \frac{2\delta\gamma^2}{u^2}\right) \right]. \quad (3)$$

2 Marginal Laplace Distribution with Irregular Grid Spacing

The following is a derivation of the marginal prior distribution for the order- k differences when grid spacing is unequal. These derivations are based on the scale-mixture representation of the Laplace distribution. These results are known to apply to the first-order and second-order models, but higher orders.

Let $u_j = \Delta^k \theta_j$ and let δ_j be a constant representing a scale factor for the distance between adjacent points when this distribution is used for the increments of a k th-order smoothing model. For convenience, subscripts on u and δ are dropped from here forward. We assume $u|\tau, \delta \sim N(0, \delta\tau^2)$ and $\tau^2|\lambda^2 \sim \text{Exp}(\lambda^2/2)$. Here $\lambda = 1/\gamma$ is the global shrinkage parameter. It follows that

$$\begin{aligned} p(u|\delta, \lambda) &= \int_0^\infty \frac{\lambda^2}{2} \exp\left(-\frac{\tau^2 \lambda^2}{2}\right) \frac{1}{\sqrt{2\pi\delta\tau^2}} \exp\left(-\frac{u^2}{2\delta\tau^2}\right) d\tau^2 \\ &= A \int_0^\infty \frac{1}{\tau} \exp\left(-\frac{\tau^2 \lambda^2}{2} - \frac{u^2}{2\delta\tau^2}\right) d\tau^2, \end{aligned}$$

where $A = \frac{\lambda^2}{2\sqrt{2\pi\delta}}$. Now we make the substitution $\omega = 1/\tau^2$, which gives $d\tau^2 = -\omega^{-2}d\omega$, and the marginal density for u becomes

$$\begin{aligned} p(u|\delta, \lambda) &= A \int_0^\infty \omega^{-3/2} \exp\left\{-\frac{\lambda^2}{2\omega} - \frac{u^2\omega}{2\delta}\right\} d\omega \\ &= A \int_0^\infty \omega^{-3/2} \exp\left\{-\frac{u^2\omega}{2\delta} - \frac{\lambda^2}{2\omega} + \frac{\lambda|u|}{\delta^{1/2}} - \frac{\lambda|u|}{\delta^{1/2}}\right\} d\omega \\ &= A \int_0^\infty \omega^{-3/2} \exp\left\{-\frac{|u|^2}{2\delta\omega} \left(\omega^2 - \frac{2\delta^{1/2}\omega\lambda}{|u|} + \frac{\delta\lambda^2}{|u|^2}\right) - \frac{\lambda|u|}{\delta^{1/2}}\right\} d\omega \\ &= \frac{\lambda}{2\sqrt{\delta}} \exp\left\{-\frac{\lambda|u|}{\sqrt{\delta}}\right\} \int_0^\infty \frac{\lambda}{\sqrt{2\pi}\omega^{3/2}} \exp\left\{-\frac{\lambda^2}{2\delta\omega(\lambda^2/|u|^2)} \left(\omega - \frac{\sqrt{\delta}\lambda}{|u|}\right)^2\right\} d\omega \\ &= \frac{\lambda}{2\sqrt{\delta}} \exp\left\{-\frac{\lambda|u|}{\sqrt{\delta}}\right\}, \end{aligned}$$

where the last line follows from the fact that the integrand in the second-to-last line is the pdf of an inverse-Gaussian distribution with mean parameter $\mu = \sqrt{\delta}\lambda/|u|$ and shape parameter $\lambda = \lambda^2$. The result is the pdf of a Laplace distribution with mean zero and scale parameter $\lambda/\sqrt{\delta}$. Note that the variance of the Laplace distribution is $2\delta/\lambda^2$, which implies that the grid spacing δ scales the variance of the increments u .

3 Data Example with Irregular Grid

We apply the SPMRF models to a data set with a continuous covariate. The response data are rent per square meter of floor space in Munich, Germany, and the covariate is the floor space in square meters. These data were analysed by Rue and Held (2005) using a second-order GMRF with irregular spacing. Here we apply a second-order GMRF and SPMRF models using the methods described in Section 2.4 of the main text.

Let x represent the floor space measurements, and let $x_1 < x_2 < \dots < x_n$ be the ordered set of unique floor space measurements. Further, let $\delta_j = x_{j+1} - x_j$ be the distance between adjacent floor space measurements. The marginal prior distributions for the second-order differences were $\Delta^2\theta_j \sim \text{N}(0, d_j\gamma^2)$ for the normal prior, $\Delta^2\theta_j \sim \text{Laplace}(d_j^{1/2}\gamma)$ for the Laplace, and $\Delta^2\theta_j \sim \text{HS}(d_j^{1/2}\gamma)$ for the horseshoe, where

$$\Delta^2\theta_j = \theta(x_{j+2}) - \left(1 + \frac{\delta_{j+1}}{\delta_j}\right)\theta(x_{j+1}) + \frac{\delta_{j+1}}{\delta_j}\theta(x_j),$$

and

$$d_j = \frac{\delta_{j+1}^2(\delta_j + \delta_{j+1})}{2}.$$

Using methods described in Section 4.1 of the main text and Section 4 of the Supplementary Material, we calculated the value of the hyperparameter for the global scale parameter to be $\zeta = 0.00094$, so $\gamma \sim \text{C}^+(0, 0.00094)$ for all models. The results are shown in Figure 1.

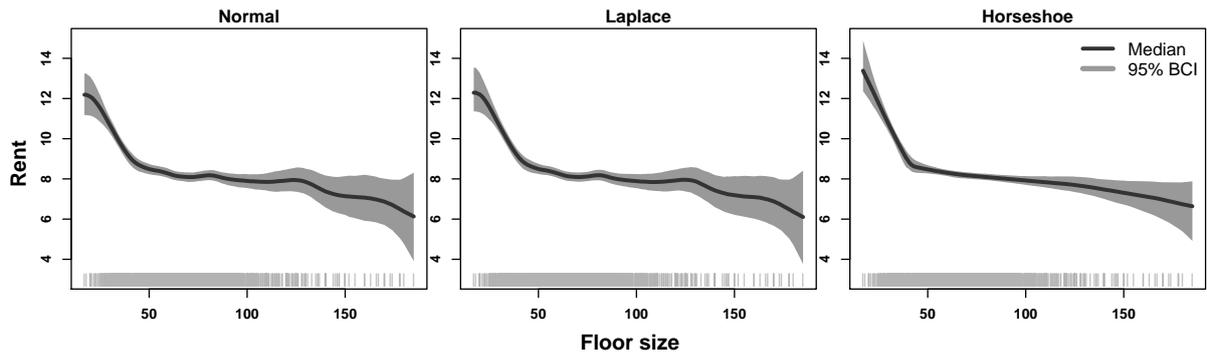


Figure 1: Results for models using irregular grids for Munich rent data. Posterior medians (dark line) are shown with 95% Bayesian credible intervals (BCIs). Locations of data are shown with vertical bars at the bottom of plots.

4 Additional Simulation Results

Here we display plots with simulation results for normal data with $\sigma = 1.5$ (Figure 2), Poisson data (Figure 3), and binomial data (Figure 4). Summary measures for all data types show similar patterns to each other and to those for normal data with $\sigma = 4.5$ (Figure 2 in main article).

Table 1: Mean values of performance measures across 100 simulations for normal observations ($\sigma = 1.5$) for each model and trend function type.

Function	Model	MAD	MCIW	MASV	TMASV
Constant	Normal	0.115	0.698	0.002	0.000
	Laplace	0.116	0.731	0.002	0.000
	Horseshoe	0.127	0.921	0.004	0.000
Piecewise Const.	Normal	1.040	5.479	1.647	0.606
	Laplace	0.899	3.282	1.557	0.606
	Horseshoe	0.281	1.918	0.638	0.606
Smooth	Normal	0.565	2.985	1.391	1.406
	Laplace	0.561	2.985	1.393	1.406
	Horseshoe	0.565	2.946	1.414	1.406
Varying Smooth	Normal	0.586	3.036	0.613	0.543
	Laplace	0.550	2.898	0.592	0.543
	Horseshoe	0.438	2.228	0.558	0.543

Table 2: Mean values of performance measures across 100 simulations for Poisson observations for each model and trend function type.

Function	Model	MAD	MCIW	MASV	TMASV
Constant	Normal	0.022	0.142	0.001	0.000
	Laplace	0.023	0.149	0.001	0.000
	Horseshoe	0.025	0.167	0.001	0.000
Piecewise Const.	Normal	0.109	0.557	0.077	0.030
	Laplace	0.092	0.529	0.064	0.030
	Horseshoe	0.051	0.334	0.036	0.030
Smooth	Normal	0.078	0.379	0.072	0.079
	Laplace	0.078	0.380	0.072	0.079
	Horseshoe	0.079	0.382	0.073	0.079
Varying Smooth	Normal	0.067	0.296	0.020	0.023
	Laplace	0.066	0.295	0.020	0.023
	Horseshoe	0.058	0.277	0.020	0.023

Table 3: Mean values of performance measures across 100 simulations for binomial observations for each model and trend function type.

Function	Model	MAD	MCIW	MASV	TMASV
Constant	Normal	0.042	0.249	0.001	0.000
	Laplace	0.043	0.262	0.001	0.000
	Horseshoe	0.047	0.311	0.002	0.000
Piecewise Const.	Normal	0.229	1.191	0.166	0.066
	Laplace	0.193	1.126	0.137	0.066
	Horseshoe	0.108	0.690	0.076	0.066
Smooth	Normal	0.139	0.733	0.110	0.117
	Laplace	0.139	0.735	0.111	0.117
	Horseshoe	0.143	0.740	0.113	0.117
Varying Smooth	Normal	0.188	0.730	0.056	0.068
	Laplace	0.183	0.726	0.056	0.068
	Horseshoe	0.149	0.676	0.058	0.068

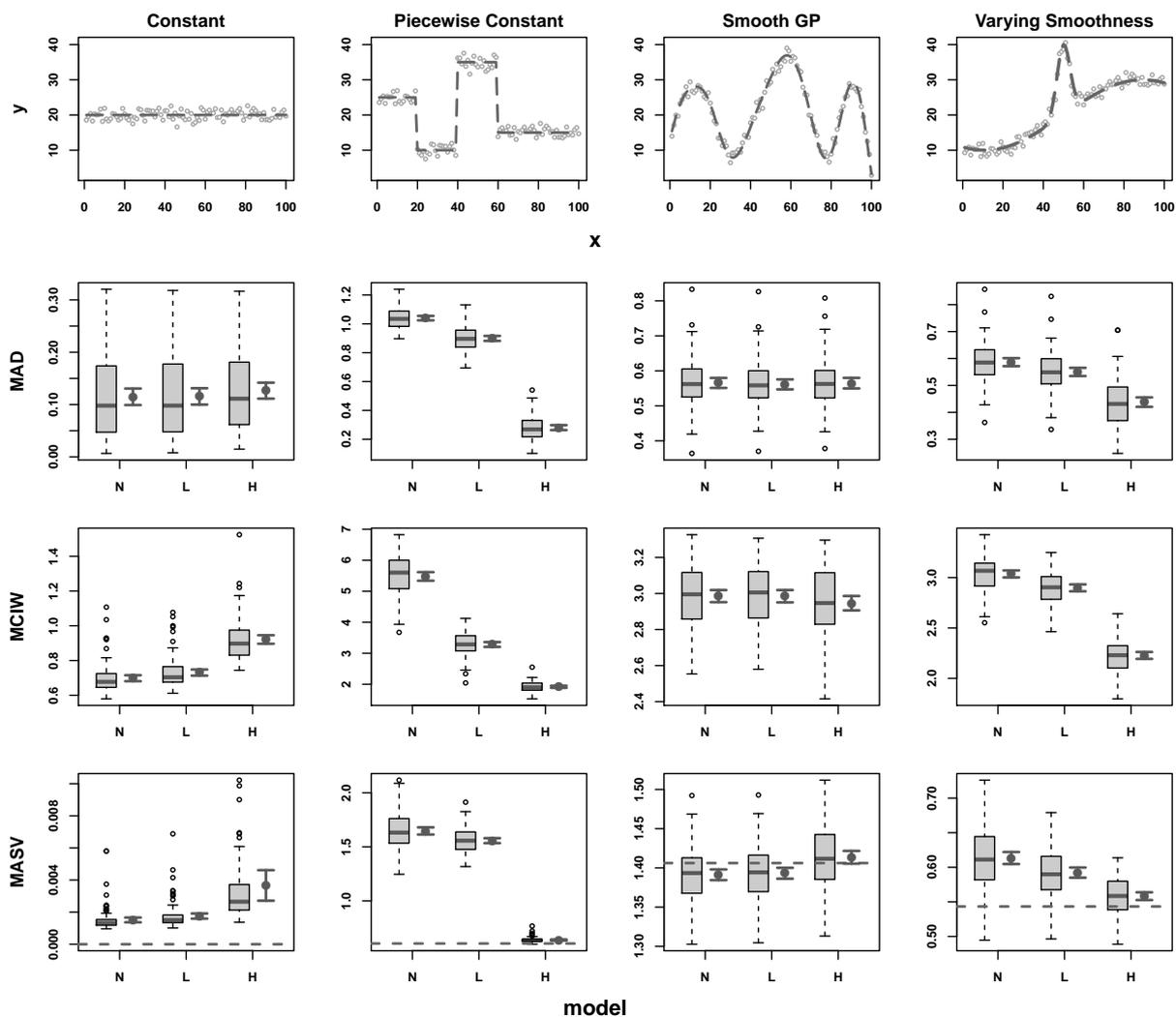


Figure 2: Functions used in simulations and simulation results by model (N=normal, L=Laplace, H=horseshoe) and function type for normally distributed data with $\sigma = 1.5$. Top row shows true functions (dashed lines) with example simulated data. Remaining rows show mean absolute deviation (MAD), mean credible interval width (MCIW), and mean absolute sequential variation (MASV). Horizontal dashed line in plots on bottom row is the true mean absolute sequential variation (TMASV). Shown for each model are standard boxplots of simulation results (left) and mean values with 95% frequentist confidence intervals (right).

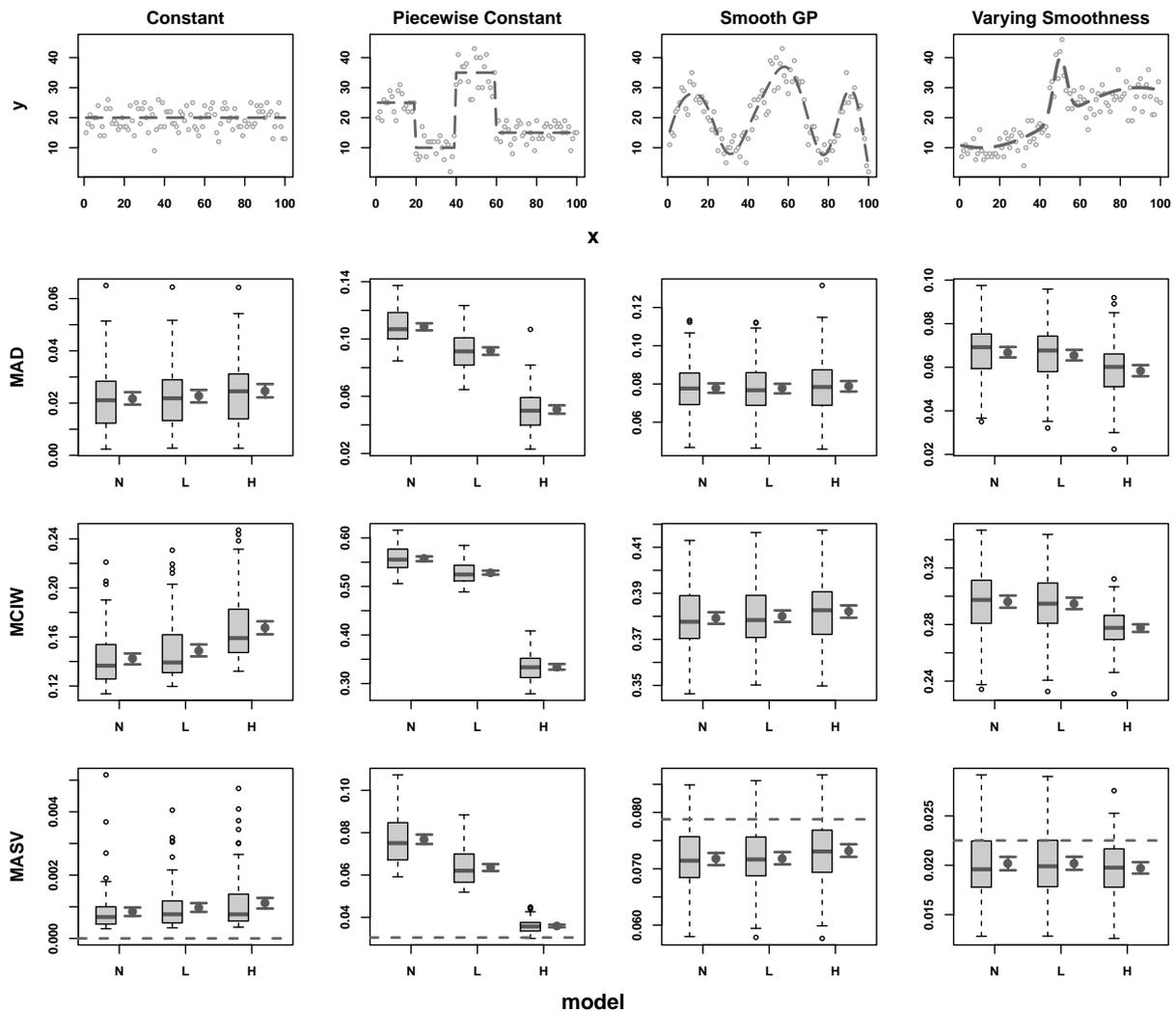


Figure 3: Functions used in simulations and simulation results by model (N=normal, L=Laplace, H=horseshoe) and function type for Poisson distributed data. Top row shows true functions (dashed lines) with example simulated data. Remaining rows show mean absolute deviation (MAD), mean credible interval width (MCIW), and mean absolute sequential variation (MASV). Horizontal dashed line in plots on bottom row is the true mean absolute sequential variation (TMA SV). Shown for each model are standard boxplots of simulation results (left) and mean values with 95 % frequentist confidence intervals (right).

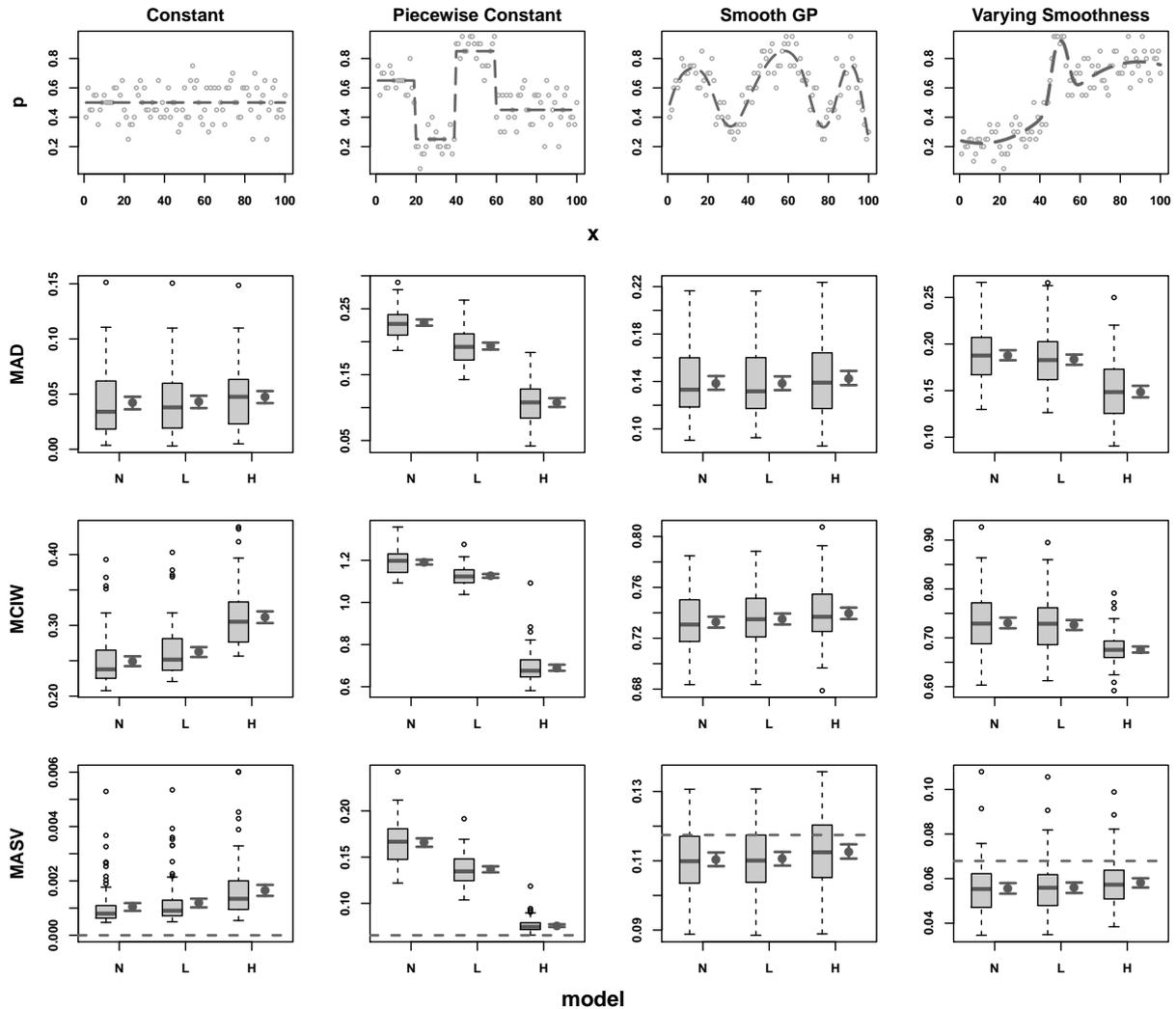


Figure 4: Functions used in simulations and simulation results by model (N=normal, L=Laplace, H=horseshoe) and function type for binomial distributed data. Top row shows true functions (dashed lines) with empirical probability estimates from example simulated data. Remaining rows show mean absolute deviation (MAD), mean credible interval width (MCIW), and mean absolute sequential variation (MASV). Horizontal dashed line in plots on bottom row is the true mean absolute sequential variation (TMASV). Shown for each model are standard boxplots of simulation results (left) and mean values with 95% frequentist confidence intervals (right).

There is an analytical form for the covariance matrix for the second-order model, but it suffices here to know that the form of the marginal variances is:

$$\Sigma_{2,ii} = \omega^2 + \frac{i(i-1)(2i-1)}{6}\gamma^2. \quad (8)$$

Note that if we were using an intrinsic GMRF model, we would assume that ω^2 is infinite, which would result in a covariance matrix of rank $n - k$. Following Sørbye and Rue (2014) we would then use the generalized inverse of the precision matrix to calculate the marginal variances.

In practice we use the variance of the data (transformed data if the θ parameters are on a transformed scale) as an estimate of ω^2 . Although this is using the data twice, this offers a reasonable constraint on the marginal variances of the θ s.

We want to set an upper bound U on the average marginal standard deviation of θ_i , such that $\Pr(\sigma_\gamma(\theta_i) > U) = \alpha$, where α is some small probability. Using the cumulative probability function for a half-Cauchy distribution, we can find a value of ζ for a given value of $\sigma_{\text{ref}}(\boldsymbol{\theta})$ specific to a model of interest and given common values of U and α by:

$$\zeta = \frac{U}{\sigma_{\text{ref}}(\boldsymbol{\theta}) \tan\left(\frac{\pi}{2}(1 - \alpha)\right)}. \quad (9)$$

It may be useful to note here that the median of a half-Cauchy distribution is equal to its scale parameter, since the median may be a more intuitive measure of the effect of different values of ζ .

For our data examples in the main text, we let U be the estimated standard deviation of the data on the appropriate scale. We know that the marginal variances of the θ s should not exceed the variance in the observed data, on average. We set $\alpha = 0.05$ as the probability of the average marginal standard deviation exceeding U . For the coal mining example in the main text, we found an estimate of the variance of the data on the log scale by $\sum_{i=1}^n \ln(y_i + 0.5)/(n - 1)$, where y_i is the observed count at time $i = 1, \dots, n$. For the Tokyo rain example, we estimated the variance of the data on the logit scale as $\sum_{i=1}^n \text{logit}((y_i + q_i)/m_i)/(n - 1)$, where y_i is the number of years with rain on day i out of m_i possible years, and $q_i = 0.005I_{y_i=0} - 0.005I_{y_i=1} + 0I_{y_i \notin \{0,1\}}$, where I is an indicator function.

Suppose we have calculated ζ_{o1} , the hyperparameter for a first-order model given the corresponding average marginal standard deviation $\sigma_{\text{ref}}(\boldsymbol{\theta}_{o1})$ using Equation (9). If we wish to calculate the value of ζ_{o2} for a second-order model we can simply use

$$\zeta_{o2} = \zeta_{o1} \frac{\sigma_{\text{ref}}(\boldsymbol{\theta}_{o1})}{\sigma_{\text{ref}}(\boldsymbol{\theta}_{o2})}.$$

Now suppose we have a model with n equidistant nodes and want to increase the density of the grid to kn nodes without changing the range of the grid. For a first-order model, $\text{Var}(\Delta\theta_{\text{new}}) = \frac{1}{k}\text{Var}(\Delta\theta)$, and for a second-order model $\text{Var}(\Delta^2\theta_{\text{new}}) = \frac{1}{k^3}\text{Var}(\Delta^2\theta)$ (Lindgren and Rue, 2008; Sørbye and Rue, 2014). In terms of the hyperparameter for the global smoothing prior, for the first-order model $\zeta_{o1,\text{new}} = k^{-1/2}\zeta_{o1}$, and for the second-order model $\zeta_{o2,\text{new}} = k^{-3/2}\zeta_{o2}$.

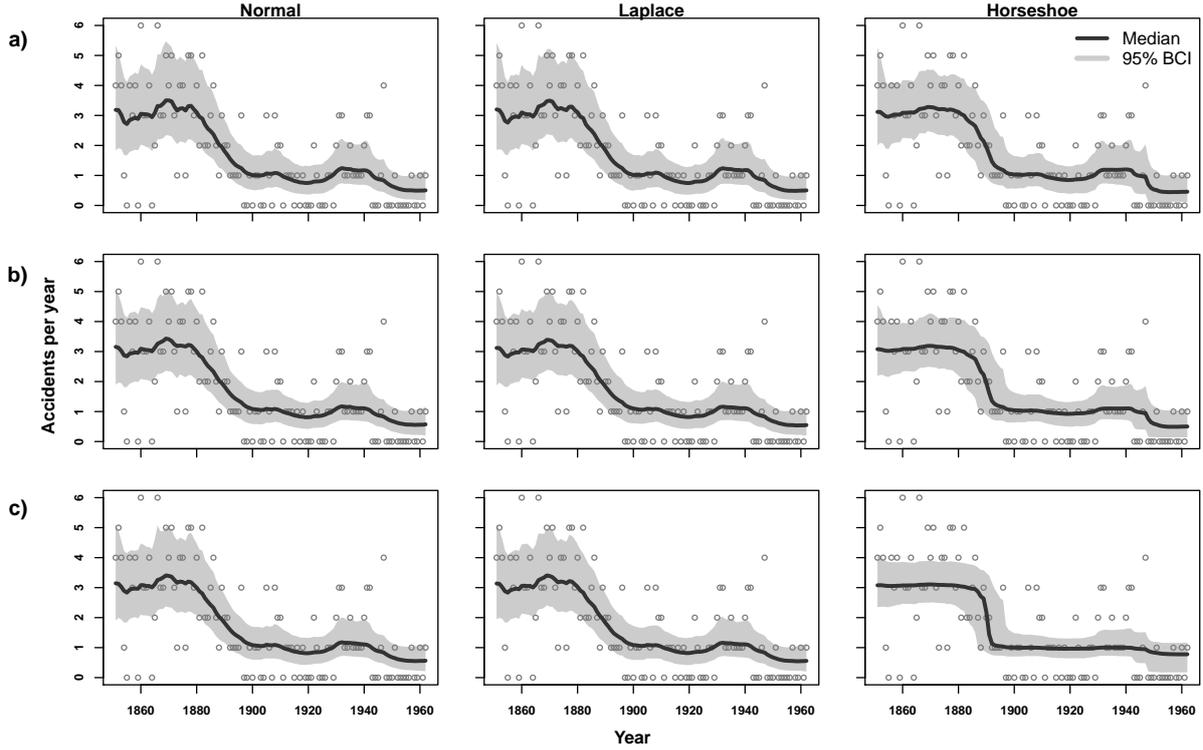


Figure 5: Models fits to coal mining accidents data by model type and value of hyperparameter for global scale: a) $\zeta = 1$, b) $\zeta = 0.01$, and c) $\zeta = 0.0001$. Posterior medians and associated 95% Bayesian credible intervals are shown along with observed data.

6 Prior Sensitivity

We tested the sensitivity of the three prior formulations (normal, Laplace, and horseshoe) to the value of the hyperparameter (ζ) which controls the scale of the distribution on the smoothing parameter γ , where $\gamma \sim C^+(0, \zeta)$. A smaller value of ζ constricts γ to be closer to zero, which in turn constricts the scales of the priors on the order- k differences. We tested three levels for the hyperparameter: a) $\zeta = 1$, b) $\zeta = 0.01$, and c) $\zeta = 0.0001$. In general, we expect noisier data sets should be more sensitive to prior settings. The coal mine disaster data offered a good test set because the observations are relatively noisy.

Clearly the horseshoe prior was the most sensitive to the level of ζ (Figure 5). In particular, the horseshoe results for $\zeta = 1$ looked more like those for the other two models in Figure 5, but when $\zeta = 0.0001$, the horseshoe produced more defined break points and straighter lines with narrower BCIs compared to the results with $\zeta = 0.01$.

7 Computational Efficiency

To compare SPMRF and GMRF models' computational efficiency, we calculated the effective number of posterior samples per second of computation time (ESSps) for different model formulations and data configurations. We used the scenario with a piecewise-constant expected value from our main simulations (Section 3) to test the effect of model type, model order, and number of grid cells (n) on the ESS per second of sampling time. Here sampling time is defined as the total run time minus the time spent in the adaptation (warm-up or burn-in) phase, where

time is measured in seconds of CPU time. We also calculated the ESSps for the coal mining and Tokyo rainfall examples.

There were three simulated scenarios with piecewise-constant trend: 1) order-1 with $n = 100$ observations and grid points (one observation per grid point), 2) order-1 with $n = 200$, and 3) order-2 with $n = 100$. The observations in these scenarios were normally distributed with standard deviation $\sigma = 4.5$. For each of these scenarios we ran 4 independent chains each with 1,000 iterations of burn-in and 2,500 iterations post burn-in thinned at every 5 for a total of 2,000 retained posterior samples combined across chains. The maximum ESS would therefore be 2,000 for these scenarios. The chains were run in sequence so that the total time (TCPU) sampling times (SCPU) times are the respective cumulative times across 4 chains. For the coal mining and Tokyo rainfall examples we used the same settings for number of iterations and thinning as was used in the main text (Section 4). We calculated effective sample size using the methods described in the documentation for `stan` (Stan Development Team, 2015).

Table 4: Measures of computational efficiency for each model type (Normal (N), Laplace (L), or Horseshoe (H)) for three simulated data scenarios and two real data examples. Model order and number of parameters (p) are shown. The total CPU time (TCPU: includes adaptive phase) and sampling CPU (SCPU) time are in seconds. The minimum and mean effective sample sizes per second (ESSps) of SCPU are also shown.

Scenario	Model	Order	p	TCPU	SCPU	Min. ESSps	Mean ESSps
Piece. Const. ($n = 100$)	N	1	102	74	52	25.14	36.79
	L	1	201	422	290	5.13	6.58
	H	1	201	1,228	897	0.80	2.12
Piece. Const. ($n = 200$)	N	1	202	198	141	10.52	13.82
	L	1	401	1,195	794	1.86	2.44
	H	1	401	2,709	1,898	0.19	1.00
Piece. Const. ($n = 100$)	N	2	102	797	592	2.61	3.26
	L	2	200	3,916	2,770	0.52	0.68
	H	2	200	4,822	3,473	0.12	0.37
Coal Mining	N	1	113	42	37	121.00	133.48
	L	1	224	228	200	20.07	24.51
	H	1	224	639	580	5.57	8.20
Tokyo Rainfall	N	2	367	20,991	18,206	0.24	0.27
	L	2	731	53,304	45,629	0.09	0.11
	H	2	731	94,128	81,891	0.03	0.06

Tests indicated that doubling the number of grid points (with a single observation per grid point) resulted in approximately 60% fewer ESSps for each model formulation (63% fewer for both the Normal and Laplace and 53% for the Horseshoe), and changing from a first- to second-order model resulted in approximately 90% fewer ESSps (91% fewer for the Normal, 90% for the Laplace, and 83% for the Horseshoe). On average across the five scenarios investigated, the Laplace formulation resulted in 77% fewer ESSps compared to the Normal formulation,

and the Horseshoe resulted in 90% fewer ESSps. In terms of sampling times, the Laplace formulations on average took 4.8 times longer to achieve the same number of effective samples as the Normal formulations (range: 2.5 to 5.7 times longer), and the Horseshoe formulations took an average of 12.2 times longer than the Normal (range: 4.6 to 17.4).

References

Carvalho, C. M., Polson, N. G., and Scott, J. G. (2010). “The Horseshoe Estimator for Sparse Signals.” *Biometrika*, 97(2): 465–480.

Lindgren, F. and Rue, H. (2008). “On the Second-Order Random Walk Model for Irregular Locations.” *Scandinavian Journal of Statistics*, 35(4): 691–700.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*. CRC Press.

Sørbye, S. H. and Rue, H. (2014). “Scaling intrinsic Gaussian Markov random field priors in spatial modelling.” *Spatial Statistics*, 8: 39–51.

Stan Development Team (2015). *Stan Modeling Language Users Guide and Reference Manual, Version 2.6.2*.

URL <http://mc-stan.org/>