Quantifying the uncertainty of contour maps

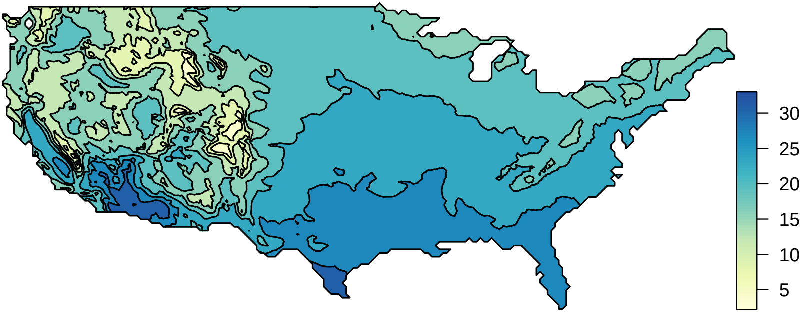
Finn Lindgren
Joint work with David Bolin

UNIVERSITY OF BATH

University of Edinburgh, 12 February 2016

Outline

Contour map for estimated US summer mean temperature



Can we trust the apparent details af the level crossings?

How many levels can we sensibly use?

Can we put a number on the statistical quality of the contour map?

Fundamental question:

What *is* the statistical interpretation of a contour map?

Spatial latent Gaussian models

Consider a simple hierarchical spatial linear model

$$\begin{aligned}\boldsymbol{\beta} &\sim \mathsf{N}(\mathbf{0}, \boldsymbol{I}\sigma_\beta^2),\\ \xi(\boldsymbol{s}) &\sim \text{Gaussian random field},\\ x(\boldsymbol{s}) &= \boldsymbol{z}(\boldsymbol{s})\boldsymbol{\beta} + \xi(\boldsymbol{s}),\\ (y_i|x) &\sim \mathsf{N}(x(\boldsymbol{s}_i), \sigma_e^2),\end{aligned}$$

where $\boldsymbol{z}(\cdot)$ are spatially indexed explanatory variables, and y_i are conditionally independent observations.

- A contour curve for a level u crossing is typically calculated as the level u crossing of $\mathsf{E}[x(\boldsymbol{s})|\boldsymbol{y}]$.
- In practice, we want to interpret it as being informative about the potential level crossings of the random field $x(\boldsymbol{s})$ itself.

Contours and excursions

- Lindgren, Rychlik (1995): *How reliable are contour curves? Confidence sets for level contours*, Bernoulli
  Regions with a single expected crossing
- Polfeldt (1999) *On the quality of contour maps*, Environmetrics
  How many contour curves should one use?
- Neither paper considered joint probabilities
- A credible contour region is a region where the field transitions from being clearly below, to being clearly above.
- Solving the problem for excursions solves it for contours.

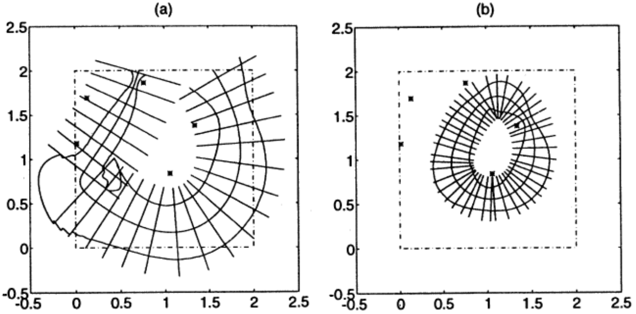
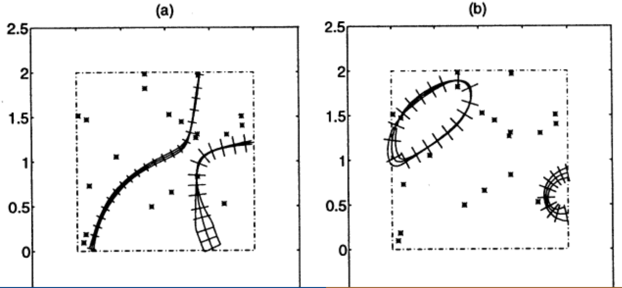


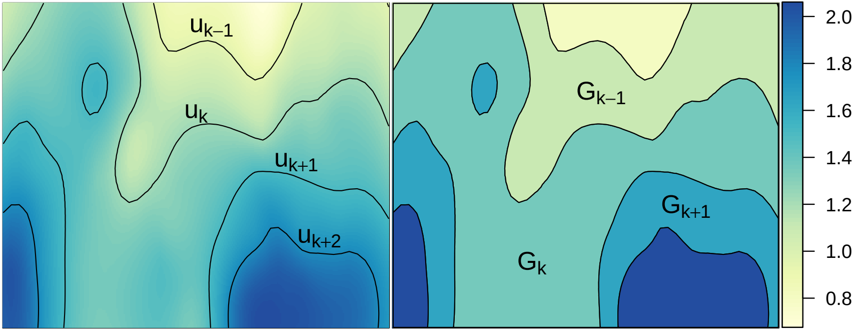
Figure 9. 50% confidence bands in Example 1 for level (a) $u = 0$, (b) $u = 2$; $n = 5$.



Level sets

Level sets

Given a function $f(s)$, $s \in \Omega$ and levels $u_1 < u_2 < \cdots < u_K$, the *level sets* are $G_k(f) = \{s;\ u_k < f(s) < u_{k+1}\}$.



Joint and marginal probabilities

Now, consider a contour map based on a point estimate $\widehat{x}(\cdot)$.

Intuitively, we might consider the joint probability

$$\mathsf{P}(u_k < x(\boldsymbol{s}) < u_{k+1},\ \text{for all } \boldsymbol{s} \in G_k(\widehat{x}) \text{ and all } k)$$

Unfortunately, this will nearly always be close to or equal to zero!

Polfeldt (1999) instead considered the marginal probability field

$$p(\boldsymbol{s}) = \mathsf{P}(u_k < x(\boldsymbol{s}) < u_{k+1} \text{ for } k \text{ such that } \boldsymbol{s} \in G_k(\widehat{x}))$$

The argument is then that if $p(\boldsymbol{s})$ is close to one in a large proportion of space, the contour map is not overconfident.

We extend this notion to alternative joint probability statements.

Contour avoiding sets and the contour map function

Contour avoiding sets

The *contour avoiding sets* $M_{\boldsymbol{u},\alpha} = (M^1_{\boldsymbol{u},\alpha}, \ldots, M^K_{\boldsymbol{u},\alpha})$ are given by

$$M_{\boldsymbol{u},\alpha} = \underset{(D_1,\ldots,D_K)}{\operatorname{argmax}} \left\{ \sum_{k=1} |D_k| : \mathsf{P}\left(\bigcap_{k=1}^{K} \{D_k \subseteq G_k(x)\} \right) \geq 1-\alpha \right\}$$

where D_k are disjoint and open sets. The joint contour avoiding set is then $C_{\boldsymbol{u},\alpha}(x) = \bigcup_{k=1}^K M^k_{\boldsymbol{u},\alpha}$.

Note: $C_{\boldsymbol{u},\alpha}(x)$ is the largest set so that with probability at least $1-\alpha$, the intuitive contour map interpretation is fulfilled for $\boldsymbol{s} \in C_{\boldsymbol{u},\alpha}(x)$.

The *contour map function* $F_{\boldsymbol{u}}(\boldsymbol{s}) = \sup\{1-\alpha;\ \boldsymbol{s} \in C_{\boldsymbol{u},\alpha}\}$ is a joint probability extension of the Polfeldt idea.

Quality measures

Let $C_{\boldsymbol{u}}(\widehat{x})$ denote a contour map based on a point estimate of x.

Three quality measures

P_0: The proportion of space where the intuitive contour map interpretation holds jointly: $P_0(x, C_{\boldsymbol{u}}(\widehat{x})) = \frac{1}{|\Omega|}\int_\Omega F_{\boldsymbol{u}}(\boldsymbol{s})\,\mathrm{d}\boldsymbol{s}$

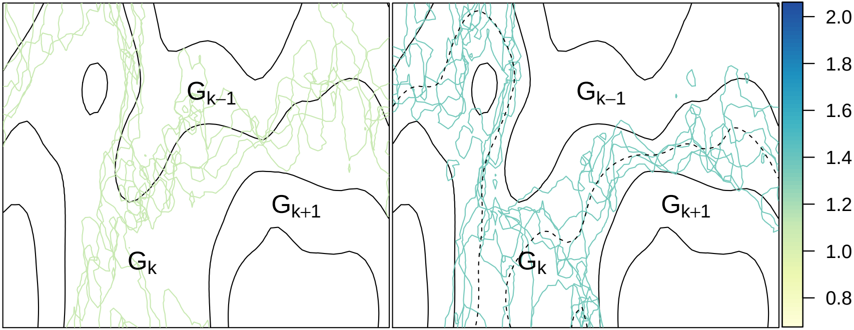
P_1: Joint credible regions for u_k crossings:

$$P_1(x, C_{\boldsymbol{u}}(\widehat{x})) = \mathsf{P}\left(\cap_k\{x(\boldsymbol{s}) < u_k \text{ where } \widehat{x}(\boldsymbol{s}) < u_{k-1}\}\cap \{x(\boldsymbol{s}) > u_k \text{ where } \widehat{x}(\boldsymbol{s}) > u_{k+1}\}\right)$$

P_2: Joint credible regions for $u_k^e = \frac{u_k+u_{k+1}}{2}$ crossings:

$$P_2(x, C_{\boldsymbol{u}}(\widehat{x})) = \mathsf{P}\left(\cap_k\{x(\boldsymbol{s}) < u_k^e \text{ where } \widehat{x}(\boldsymbol{s}) < u_k\}\cap \{x(\boldsymbol{s}) > u_k^e \text{ where } \widehat{x}(\boldsymbol{s}) > u_{k+1}\}\right)$$

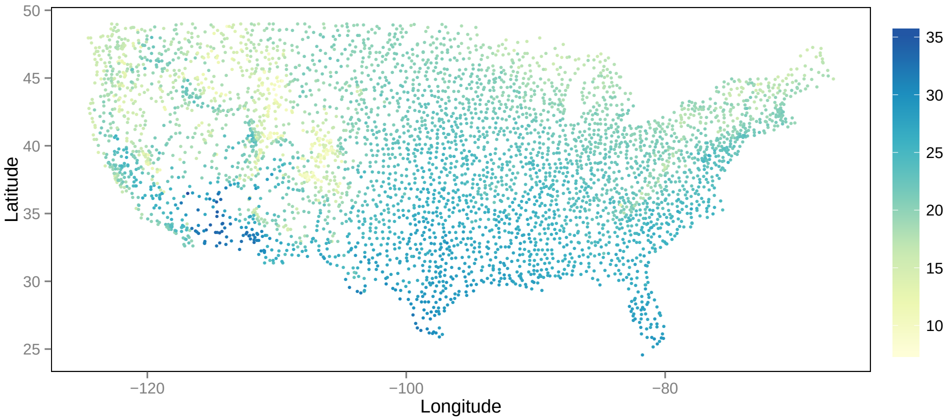
Interpretation of P_1 and P_2



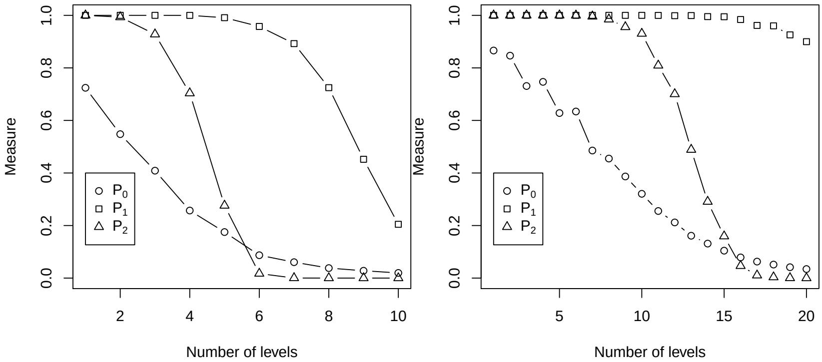
Five realisations of contour curves from the posterior distribution for x are shown.

Note the fundamental difference in smoothness between the contours of $\widehat{x}$ and x!

Mean summer temperature measurements for 1997



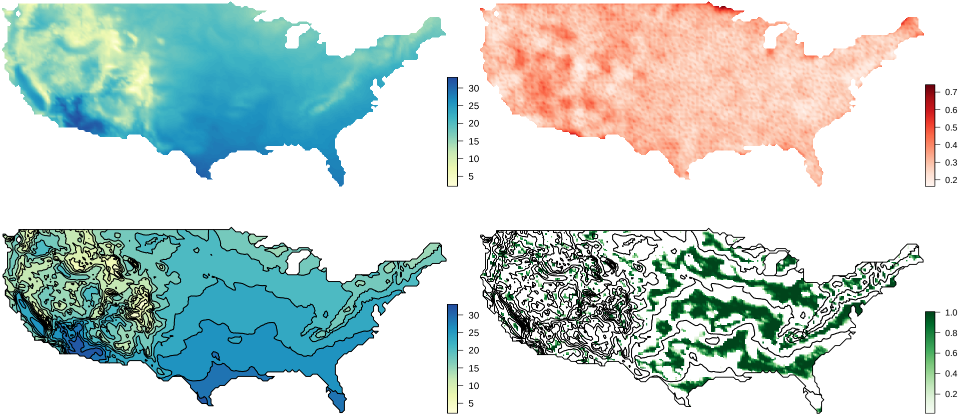
Contour map quality for different K and different models



The spatial predictions are more uncertain in a model without spatial explanatory variables (left) than in a model using elevation (right).

P_1 consistently admits about double the number of contour levels in comparison with P_2, as expected from the probabilistic interpretations.

Posterior mean, s.d., contour map, and F_u, for $K = 10$



Contour map quality measures: $P_0 = 0.38$ and $P_2 = 0.94$

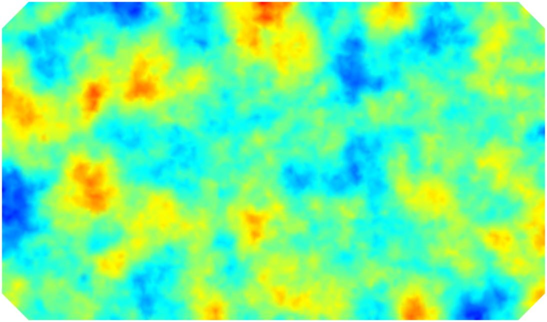
References

- David Bolin and Finn Lindgren (2015): Excursion and contour uncertainty regions for latent Gaussian models, *JRRS Series B*, 77(1):85–106
- David Bolin and Finn Lindgren (2016): Quantifying the uncertainty of contour maps, in review. `http://arxiv.org/abs/1507.01778`
- David Bolin and Finn Lindgren (2013–2016): R CRAN package `excursions`
  ```
  contourmap(mu=expectation, Q=precision)
  contourmap.inla(result.inla)
  continuous(..., geometry)
  ```
- Lindgren, F., Rue, H. and Lindström, J. (2011): An explicit link between Gaussian fields and Gaussian Markov eandom fields: the stochastic partial differential equation approach (with discussion); *JRSS Series B*, 73(4):423–498

GMRFs based on SPDEs (Lindgren et al., 2011)

GMRF representations of SPDEs can be constructed for oscillating, anisotropic, non-stationary, non-separable spatio-temporal, and multivariate fields on manifolds.

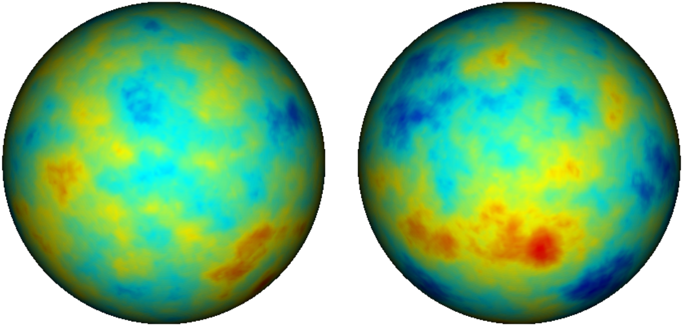
$$(\kappa^2 - \Delta)(\tau\, x(\boldsymbol{u})) = \mathcal{W}(\boldsymbol{u}), \quad \boldsymbol{u} \in \mathbb{R}^d$$



GMRFs based on SPDEs (Lindgren et al., 2011)

GMRF representations of SPDEs can be constructed for oscillating, anisotropic, non-stationary, non-separable spatio-temporal, and multivariate fields on manifolds.

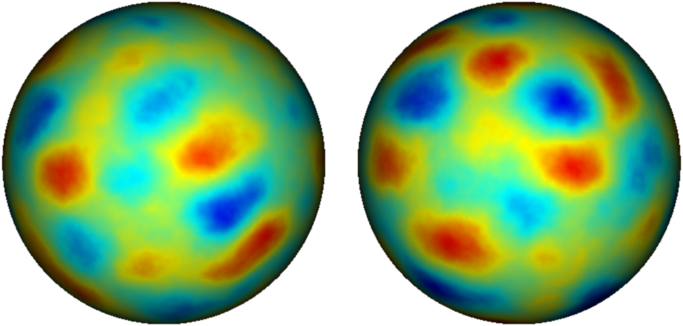
$$(\kappa^2 - \Delta)(\tau\, x(\boldsymbol{u})) = \mathcal{W}(\boldsymbol{u}), \quad \boldsymbol{u} \in \Omega$$



GMRFs based on SPDEs (Lindgren et al., 2011)

GMRF representations of SPDEs can be constructed for oscillating, anisotropic, non-stationary, non-separable spatio-temporal, and multivariate fields on manifolds.

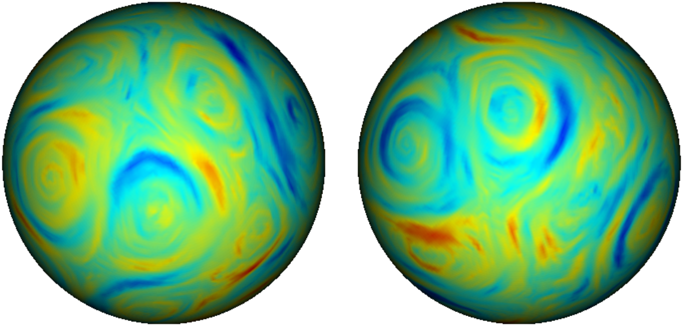
$$(\kappa^2 e^{i\pi\theta} - \Delta)(\tau x(\boldsymbol{u})) = \mathcal{W}(\boldsymbol{u}), \quad \boldsymbol{u} \in \Omega$$



GMRFs based on SPDEs (Lindgren et al., 2011)

GMRF representations of SPDEs can be constructed for oscillating, anisotropic, non-stationary, non-separable spatio-temporal, and multivariate fields on manifolds.

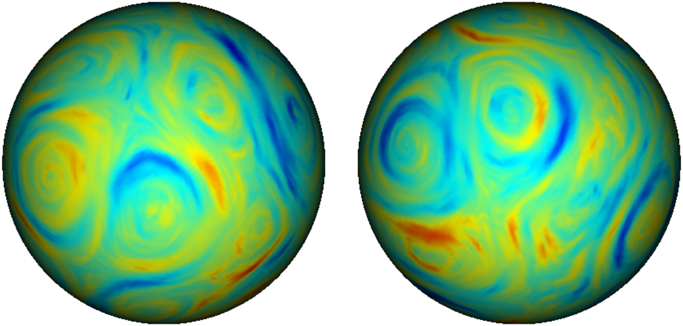
$$(\kappa_{\boldsymbol{u}}^2 + \nabla \cdot \boldsymbol{m}_{\boldsymbol{u}} - \nabla \cdot \boldsymbol{M}_{\boldsymbol{u}} \nabla)(\tau_{\boldsymbol{u}} x(\boldsymbol{u})) = \mathcal{W}(\boldsymbol{u}), \quad \boldsymbol{u} \in \Omega$$



GMRFs based on SPDEs (Lindgren et al., 2011)

GMRF representations of SPDEs can be constructed for oscillating, anisotropic, non-stationary, non-separable spatio-temporal, and multivariate fields on manifolds.

$$\left(\frac{\partial}{\partial t} + \kappa^2_{\boldsymbol{u},t} + \nabla \cdot \boldsymbol{m}_{\boldsymbol{u},t} - \nabla \cdot \boldsymbol{M}_{\boldsymbol{u},t} \nabla\right)(\tau_{\boldsymbol{u},t} x(\boldsymbol{u},t)) = \mathcal{E}(\boldsymbol{u},t), \quad (\boldsymbol{u},t) \in \Omega \times \mathbb{R}$$



A sequential Monte-Carlo algorithm

- A GMRF can be viewed as a non-homogeneous AR-process defined backwards in the indices of $\boldsymbol{x} \sim \mathsf{N}(\boldsymbol{\mu}, \boldsymbol{Q}^{-1})$.
- Let $\boldsymbol{L}$ be the Cholesky factor in $\boldsymbol{Q} = \boldsymbol{L}\boldsymbol{L}^\top$. Then

$$x_i | x_{i+1}, \ldots, x_n \sim \mathsf{N}\left(\mu_i - \frac{1}{L_{ii}} \sum_{j=i+1}^{n} L_{ji}(x_j - \mu_j), L_{ii}^{-2}\right)$$

- Denote the integral of the last $d - i$ components as I_i,

$$I_i = \int_{a_i}^{b_i} \pi(x_i | x_{i+1:d}) \cdots \int_{a_{d-1}}^{b_{d-1}} \pi(x_{d-1} | x_d) \int_{a_d}^{b_d} \pi(x_d) \, \mathrm{d}x,$$

- $x_i | x_{i+1:d}$ only depends on the elements in $x_{\mathcal{N}_i \cap \{i+1:d\}}$.
- Estimate the integrals using sequential importance sampling.
- In each step x_j is sampled from the truncated Gaussian distribution $1(a_j < x_j < b_j)\pi(x_j | x_{j+1:d})$.
- The importance weights can be updated recursively.

Extension to a latent Gaussian setting

- Assuming that $\pi(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})$ is, or can be approximated as, Gaussian, there are several ways to calculate the excursion probabilities, one of which is

Numerical integration

Numerically approximate the excursion probability by approximating the posterior integral as

$$\mathsf{P}(\boldsymbol{a} < \boldsymbol{x} < \boldsymbol{b}) = \mathsf{E}[\mathsf{P}(\boldsymbol{a} < \boldsymbol{x} < \boldsymbol{b}|\boldsymbol{\theta})] \approx \sum_{i=1}^{k} w_i \mathsf{P}(\boldsymbol{a} < \boldsymbol{x} < \boldsymbol{b}|\boldsymbol{\theta}_i),$$

where the configuration $\{\boldsymbol{\theta}_i\}$ is taken from INLA and the weights w_i are chosen proportional to $\pi(\boldsymbol{\theta}_i|\boldsymbol{y})$.

- Often only a few configurations $\{\boldsymbol{\theta}_i\}$ are needed.