Quantifying the uncertainty of contour maps

Finn Lindgren
Chair of Statistics
finn.lindgren@ed.ac.uk
Joint work with David Bolin

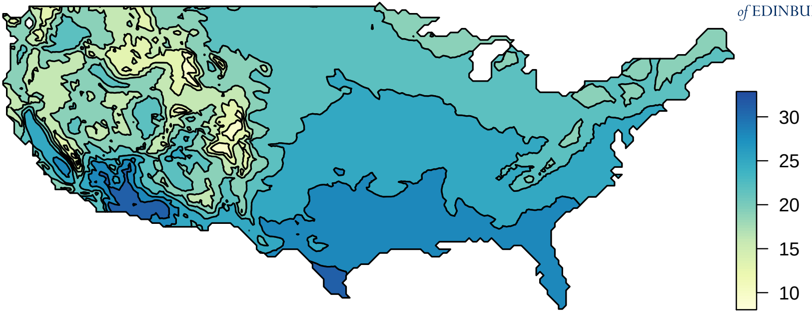
THE UNIVERSITY of EDINBURGH

Scotland

METMA IX, Montpellier, France, 14th June 2018

Contour map for US summer mean temperature

THE UNIVERSITY of EDINBURGH



- Can we trust the apparent details af the level crossings?
- How many levels should 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?
- To answer these questions we need methods for efficient calculations for random fields.

GMRFs: Gaussian Markov random fields

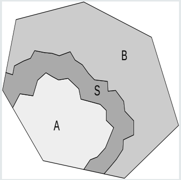
THE UNIVERSITY of EDINBURGH

Continuous domain GMRFs ((Rozanov, 1977)

If $x(\boldsymbol{s})$ is a (stationary) Gaussian random field on Ω with covariance function $R_x(\boldsymbol{s}, \boldsymbol{s}')$, it fulfills the *global Markov property*

$$\{x(\mathcal{A}) \perp x(\mathcal{B}) | x(\mathcal{S}), \text{ for all } \mathcal{AB}\text{-separating sets } \mathcal{S} \subset \Omega\}$$

if the power spectrum can be written as $1/S_x(\boldsymbol{\omega}) =$ polynomial in $\boldsymbol{\omega}$, for some polynomial order p.
Generally: Markov if the precision operator is local.



Discrete domain GMRFs

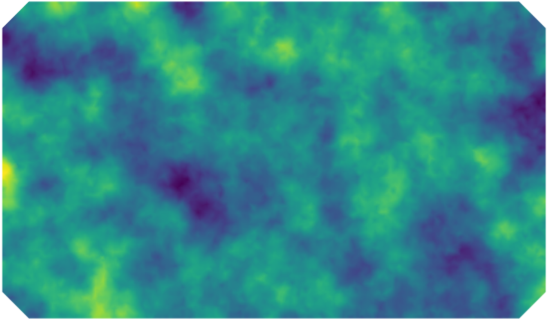
$\boldsymbol{x} = (x_1, \ldots, x_n) \sim \mathsf{N}(\boldsymbol{\mu}, \boldsymbol{Q}^{-1})$ is Markov with respect to a neighbourhood structure $\{\mathcal{N}_i, i = 1, \ldots, n\}$ if $Q_{ij} = 0$ whenever $j \neq \mathcal{N}_i \cup i$.

- Continuous domain basis representation with Markov weights: $x(\boldsymbol{s}) = \sum_{k=1}^{n} \Psi_k(\boldsymbol{s}) x_k$
- Many stochastic PDE solutions are Markov in continuous space, and can be approximated by *Markov weights on local basis functions*.

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.

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

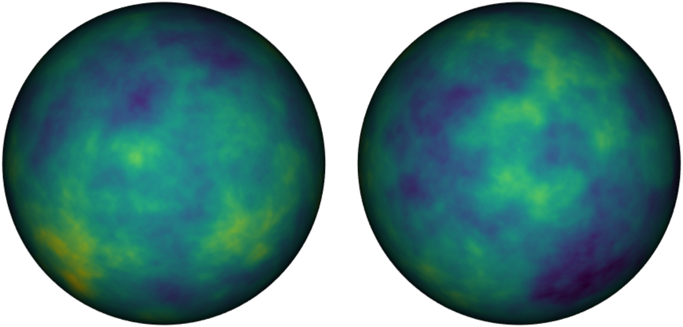


GMRFs based on SPDEs (Lindgren et al., 2011)

THE UNIVERSITY of EDINBURGH

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

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

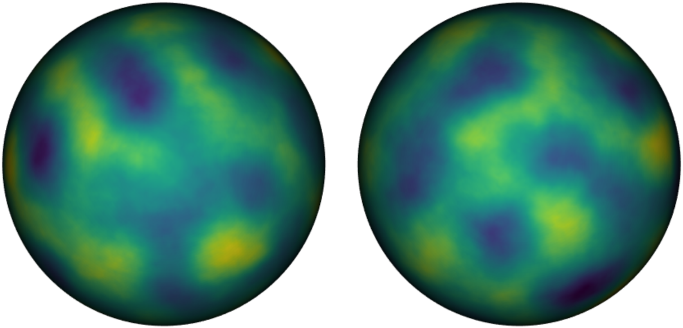


GMRFs based on SPDEs (Lindgren et al., 2011)

THE UNIVERSITY of EDINBURGH

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

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

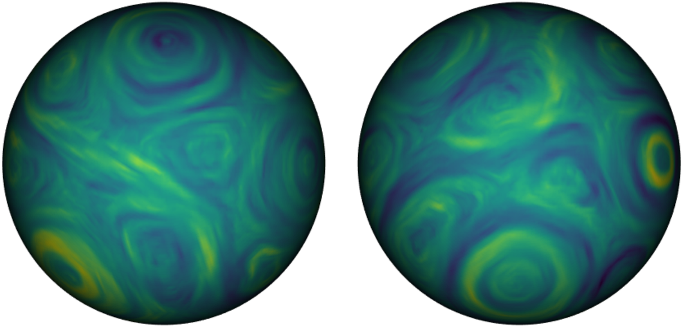


GMRFs based on SPDEs (Lindgren et al., 2011)

THE UNIVERSITY of EDINBURGH

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

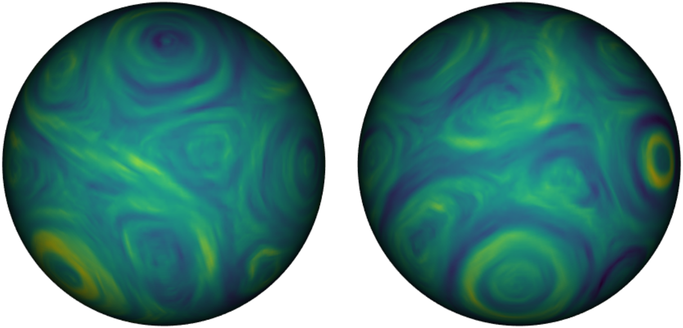
$$(\kappa_{\boldsymbol{s}}^2 + \nabla \cdot \boldsymbol{m}_{\boldsymbol{s}} - \nabla \cdot \boldsymbol{M}_{\boldsymbol{s}} \nabla)(\tau_{\boldsymbol{s}} x(\boldsymbol{s})) = \mathcal{W}(\boldsymbol{s}), \quad \boldsymbol{s} \in \Omega$$



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.

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



Spatial latent Gaussian models

THE UNIVERSITY of EDINBURGH

Consider a simple hierarchical spatial generalised linear model

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

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 $\widehat{x} = \mathsf{E}[x(\boldsymbol{s})|\boldsymbol{y}]$.
- In practice, we want to interpret it as being informative about the potential level crossings of the random field $x(\boldsymbol{s})$ itself.
- We need access to high dimensional joint probabilities in the posterior density $\pi(\boldsymbol{x}|\boldsymbol{y})$.

Posterior probabilities

THE UNIVERSITY of EDINBURGH

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

Numerical integration

Numerically approximate the excursion probability by approximating the posterior integral as

$$\mathsf{P}(\boldsymbol{a} < \boldsymbol{x} < \boldsymbol{b}|\boldsymbol{y}) = \mathsf{E}[\mathsf{P}(\boldsymbol{a} < \boldsymbol{x} < \boldsymbol{b}|\boldsymbol{y},\boldsymbol{\theta})] \approx \sum_k w_k \mathsf{P}(\boldsymbol{a} < \boldsymbol{x} < \boldsymbol{b}|\boldsymbol{y},\boldsymbol{\theta}_k),$$

where each parameter configuration $\boldsymbol{\theta}_k$ is provided by R-INLA and the weights w_k are chosen proportional to $\pi(\boldsymbol{\theta}_k|\boldsymbol{y})$.

- Often only a few configurations $\boldsymbol{\theta}_k$ are needed.
- Quantile corrections and other techniques from INLA can be added

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 $n-i$ components as I_i,

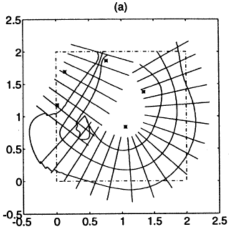
$$I_i = \int_{a_i}^{b_i} \pi(x_i | x_{i+1:n}) \cdots \int_{a_{n-1}}^{b_{n-1}} \pi(x_{n-1} | x_n) \int_{a_n}^{b_n} \pi(x_n)\, \mathrm{d}x,$$

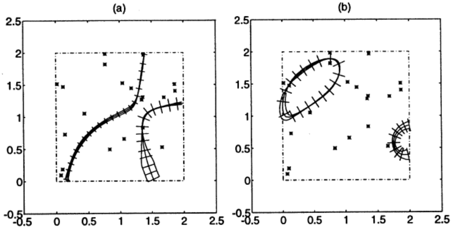
- $x_i | x_{i+1:n}$ only depends on the elements in $x_{\mathcal{N}_i \cap \{i+1:n\}}$.
- Estimate the integrals using sequential importance sampling.
- In each step x_j is sampled from the truncated Gaussian density $\propto \mathbb{I}_{\{a_j < x_j < b_j\}} \pi(x_j | x_{j+1:n})$.
- The importance weights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

Contours and excursions

THE UNIVERSITY *of* EDINBURGH

- 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.





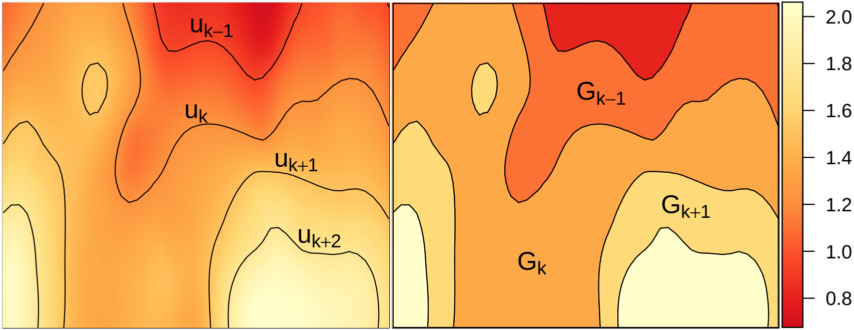
Level sets

Level sets

Given a function $f(\boldsymbol{s})$, $\boldsymbol{s} \in \Omega$ and levels

$$-\infty = u_0 < u_1 < u_2 < \cdots < u_K < u_{K+1} = +\infty,$$

the *level sets* are $G_k(f) = \{\boldsymbol{s};\ u_k < f(\boldsymbol{s}) < u_{k+1}\}$, $k = 0, \ldots, K$.



Joint and marginal probabilities

THE UNIVERSITY *of* EDINBURGH

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 1 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

THE UNIVERSITY of EDINBURGH

Contour avoiding sets

The *contour avoiding sets* $M_{\boldsymbol{u},\alpha} = (M^0_{\boldsymbol{u},\alpha}, \ldots, M^K_{\boldsymbol{u},\alpha})$ are given by

$$M_{\boldsymbol{u},\alpha} = \underset{(D_0,\ldots,D_K)}{\operatorname{argmax}} \left\{ \sum_{k=0}^{K} |D_k| : \mathsf{P}\left(\bigcap_{k=0}^{K} \{D_k \subseteq G_k(x)\} \right) \geq 1-\alpha \right\}$$

where D_k are disjoint and open sets. The joint contour avoiding set is then $C_{\boldsymbol{u},\alpha}(x) = \bigcup_{k=0}^{K} M^k_{\boldsymbol{u},\alpha}$.

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

THE UNIVERSITY of EDINBURGH

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: $P_0(x, C_{\boldsymbol{u}}(\widehat{x})) = \frac{1}{|\Omega|}\int_\Omega F_{\boldsymbol{u}}(\boldsymbol{s})\,\mathrm{d}\boldsymbol{s}$

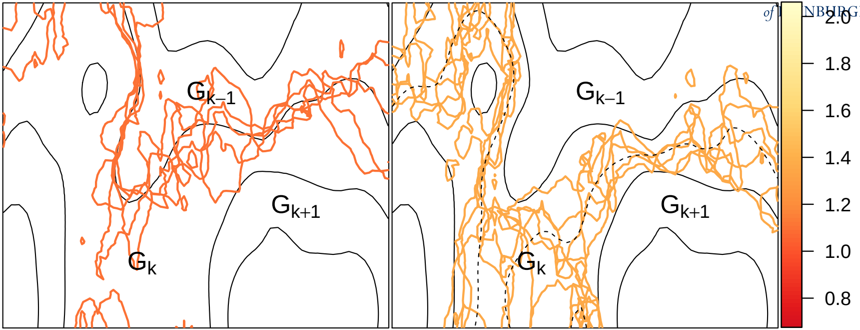
P_1: Joint credible regions for u_k crossings:

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

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

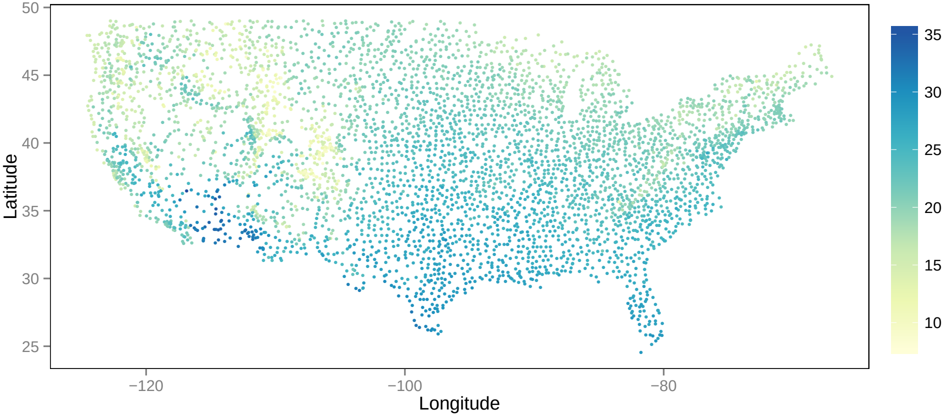


Five realisations of contour curves from the posterior distribution for x are shown.

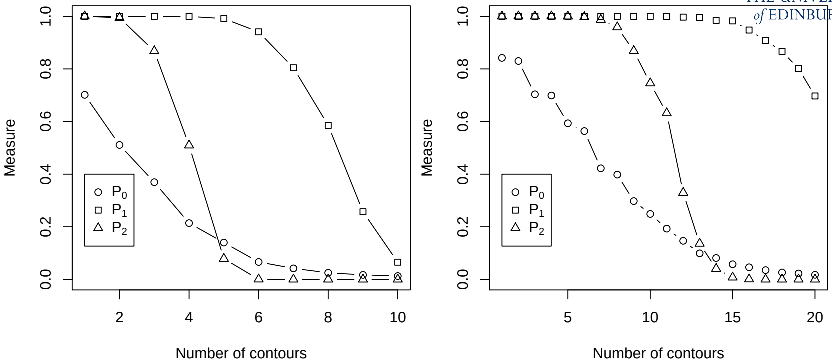
Note the fundamental difference in smoothness between the contours of $\widehat{x}$ and x!

Additional note for theorists:
The process x is not a member of its own RKHS, but $\hat{x}$ usually is.
This is a feature, not a bug.

Mean summer temperature measurements for 1997



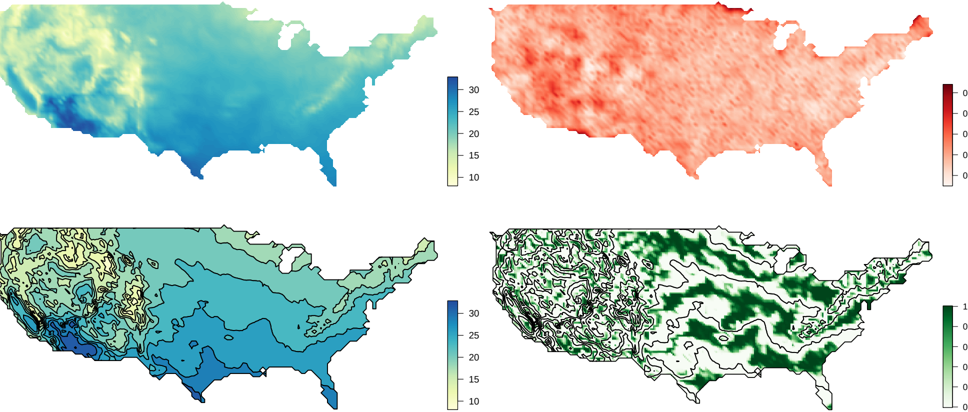
Contour map quality for different K and different models



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 $K = 8$



Contour map quality measure: $P_2 = 0.958$

Summary

THE UNIVERSITY *of* EDINBURGH

- Drawn contours are usually non-linear functions of point estimates
- Point estimate contour shapes do not match actual structure
- Recast the uncertainty problem as probabilistic excursion sets
- Excursion formulation allows discontinuities, avoiding the hypothesis testing *equal to the level* trap
- Recursive Monte Carlo integration for high dimensional probabilities
- General concept not tied to a specific computational method
- Instead of drawing too many contours, should often consider either
  - using a continuous colour scale, showing the entire point estimate, or
  - using only a specific contour of interest, e.g. "regulation air quality limit"

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, *J of Computational and Graphical Statistics*.
- David Bolin and Finn Lindgren (2018, to appear): Calculating Probabilistic Excursion Sets and Related Quantities Using `excursions`, *J of Statistical Software*. `https://arxiv.org/abs/1612.04101`
- David Bolin and Finn Lindgren: R CRAN package `excursions` (dev on `https://bitbucket.com/davidbolin/excursions`)

```
contourmap(mu = expectation, Q = precision)
contourmap.inla(result.inla) # INLA or inlabru output
continuous(..., geometry) # Interpret on continuous domain
```

- 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