(Towards) Non-stationary distributional regression methods for historical climate analysis

Finn Lindgren (finn.lindgren@ed.ac.uk)

THE UNIVERSITY *of* EDINBURGH

CMStatistics, Berlin, 2023-12-18

Integrating specialised models into general purpose software

- Revisiting the EUSTACE project; unfinished business
- Diurnal temperature range
  - Spatially and seasonally varying distributions
  - Two-stage estimation; 1) time series model, 2) spatial distribution interpolation
- `INLA`: Bayesian Generalized additive models with latent Gaussian processes
- `inlabru`: Iterated linearised INLA
- Towards joint estimation
  - Observation models with multiple predictors
  - Non-Gaussian latent models

Station observation&homogenisation model

Daily means

For station k at day t_i,

$$y_m^{k,i} = T_m(\mathbf{s}_k, t_i) + \sum_{j=1}^{J_k} H_j^k(t_i) e_m^{k,j} + \epsilon_m^{k,i},$$

where $H_j^k(t)$ are temporal step functions, $e_m^{k,j}$ are latent bias variables, and $\epsilon_m^{k,i}$ are independent measurement and discretisation errors.

The total correction term is $\widetilde{H}_m^k(t) = \sum_{j=1}^{J_k} H_j^k(t_i) e_m^{k,j}$.

Mean/max/min observations

Daily mean/max/min

For station k at day t_i,

$$
\begin{aligned}
y_m^{k,i} &= T_m(\mathbf{s}_k, t_i) + \widetilde{H}_m^k(t_i) + \epsilon_m^{k,i}, \\
y_x^{k,i} &= T_m(\mathbf{s}_k, t_i) + \frac{\exp[\widetilde{H}_r^k(t_i)]}{2} T_r(\mathbf{s}_k, t_i) + \epsilon_x^{k,i}, \\
y_n^{k,i} &= T_m(\mathbf{s}_k, t_i) - \frac{\exp[\widetilde{H}_r^k(t_i)]}{2} T_r(\mathbf{s}_k, t_i) + \epsilon_n^{k,i},
\end{aligned}
$$

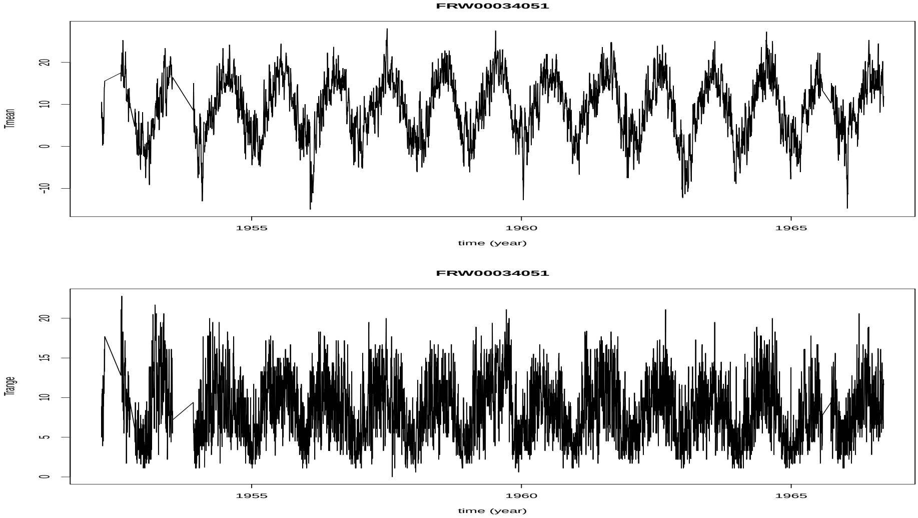
where $\widetilde{H}_\cdot^\cdot$ are the total bias correction variables for each observation.

Simplification: model the diurnal temperature range directly:

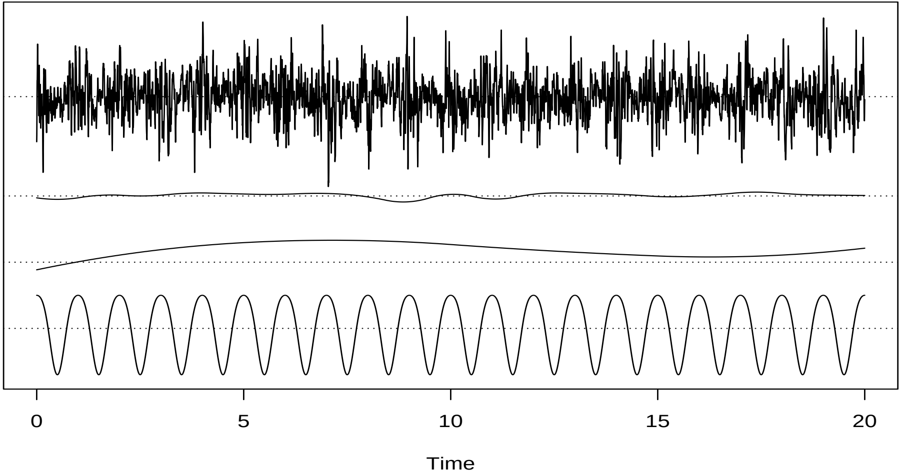
$$
y_r^{k,i} = y_x^{k,i} - y_n^{k,i} = \exp[\widetilde{H}_r^k(t_i)] T_r(\mathbf{s}_k, t_i) + \epsilon_r^{k,i}
$$

Observed data

Observed daily T_{mean} and T_{range} for station FRW00034051

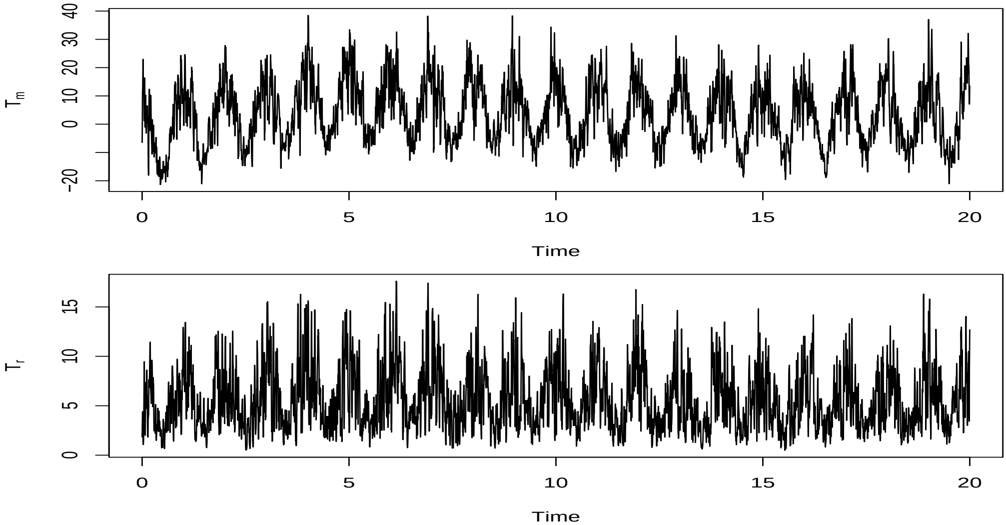


Multiscale model component samples



Combined model samples for T_m and T_r

(Proof of concept; no actual data was involved in this figure)



Modelling non-Gaussian quantities

Power tail quantile (POQ) model

The quantile function $F_{\boldsymbol{\theta}}^{-1}(p)$, $p \in [0,1]$, is defined through a quantile blend of left- and right-tailed generalised Pareto distributions:

$$f_\theta^-(p) = \begin{cases} \frac{1-(2p)^{-\theta}}{2\theta}, & \theta \neq 0, \\ \frac{1}{2}\log(2p), & \theta = 0, \end{cases}$$

$$f_\theta^+(p) = -f_\theta^-(1-p) = \begin{cases} \frac{(2(1-p))^{-\theta}-1}{2\theta}, & \theta \neq 0, \\ -\frac{1}{2}\log(2(1-p)), & \theta = 0. \end{cases}$$

$$F_{\boldsymbol{\theta}}^{-1}(p) = \theta_0 + \frac{\tau}{2}\left[(1-\gamma)f_{\theta_3}^-(p) + (1+\gamma)f_{\theta_4}^+(p)\right].$$

The parameters $\boldsymbol{\theta} = (\theta_0, \theta_1 = \log\tau, \theta_2 = \operatorname{logit}[(\gamma+1)/2], \theta_3, \theta_4)$ control the median, spread/scale, skewness, and the left and right tail shape.
This model is also known as the *five parameter lambda model* (Gilchrist, 2000).

Converting Gaussian to POQ

A POQ copula model

A spatio-temporally dependent Gaussian field $u(\mathbf{s},t)$ with expectation 0 and variance 1 can be transformed into a POQ field by

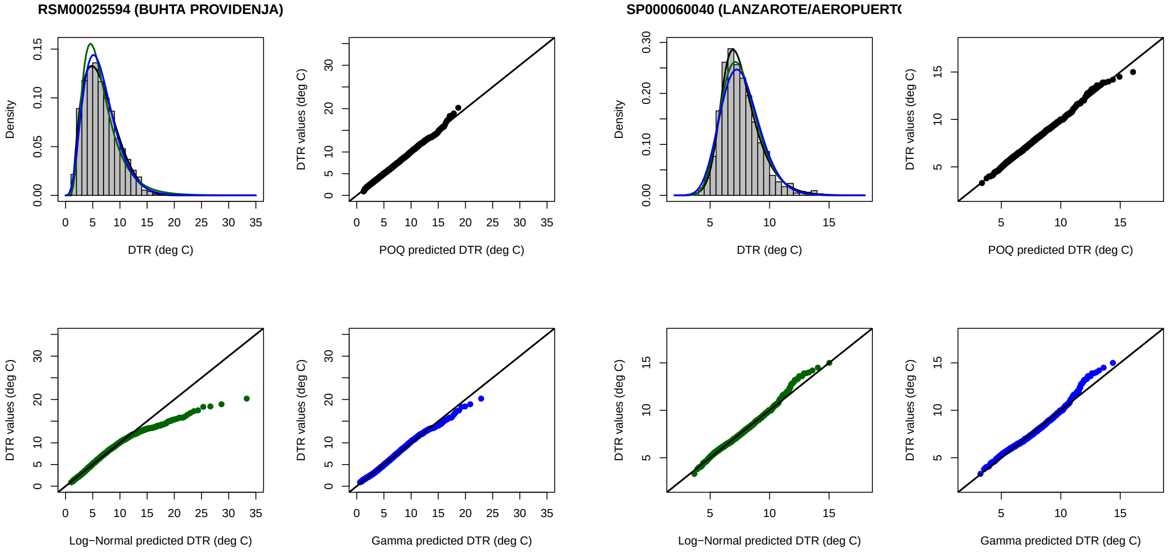
$$\widetilde{u}(\mathbf{s},t) = G^{-1}[u(\mathbf{s},t)] = F^{-1}_{\boldsymbol{\theta}(\mathbf{s},t)}[\Phi(u(\mathbf{s},t))],$$

where the parameters can vary with space and time.

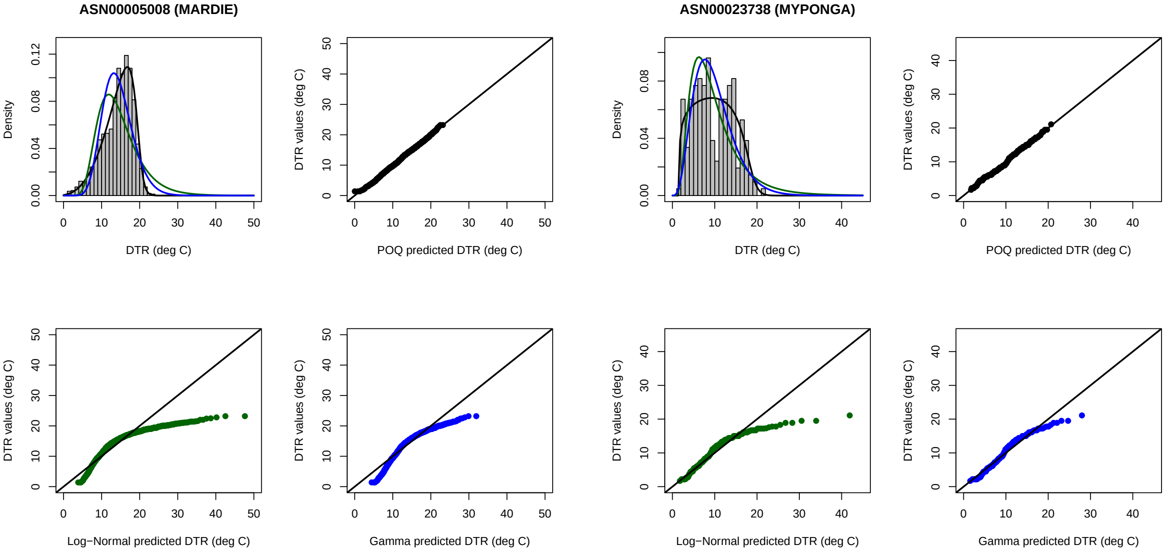
Ignoring the homogenisation model, can estimate using a two-step procedure:

1. Estimate seasonal POQ and temporal covariance parameters for separate time series
2. Apply a basic spatial-seasonal random field model for the parameters

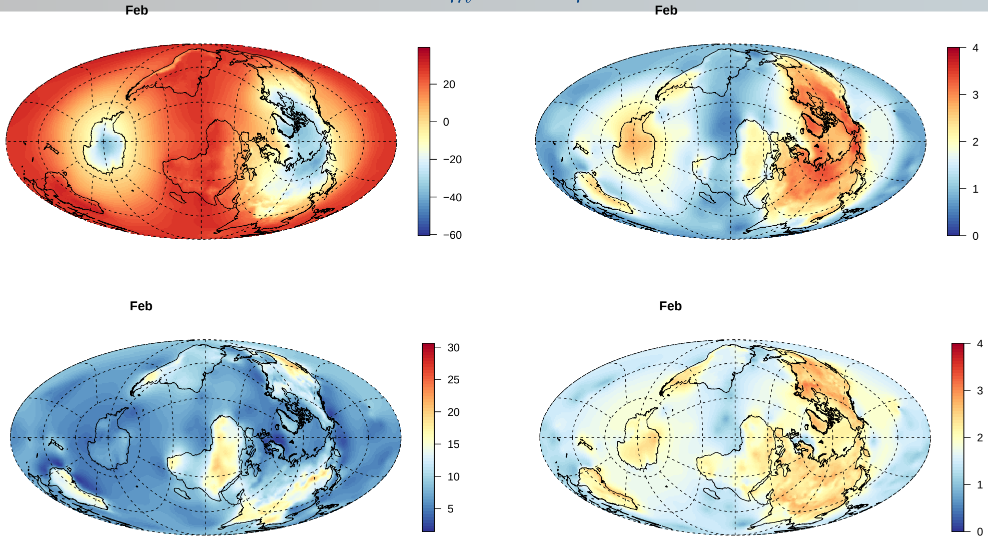
Diurnal range distributions



Diurnal range distributions



Estimates of median & scale for T_m and T_r



February climatology

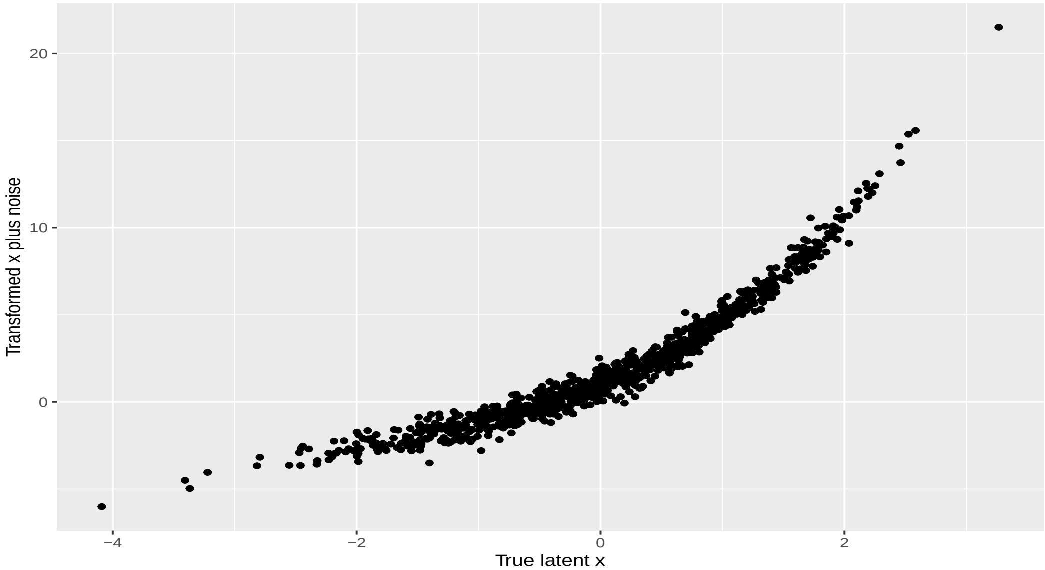
Challenges for joint estimation

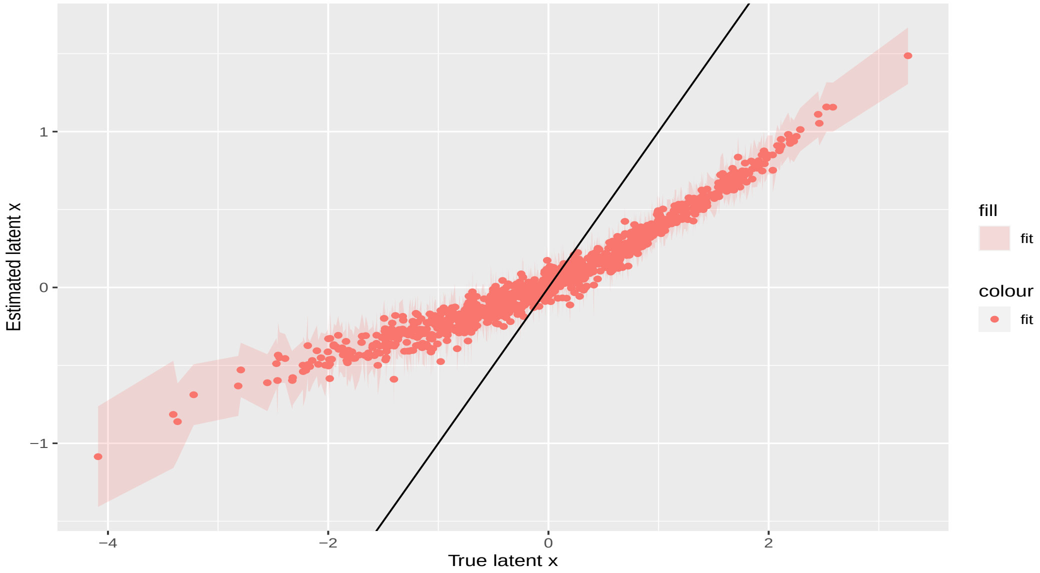
- Integrate the POQ model estimation with the homogenisation estimation
- Use the POQ transformation model as a component in a Bayesian generalised additive model (GAM)
- Use the POQ model as observation model in a Bayesian GAM

Latent Gaussian models and INLA (with simplified details):

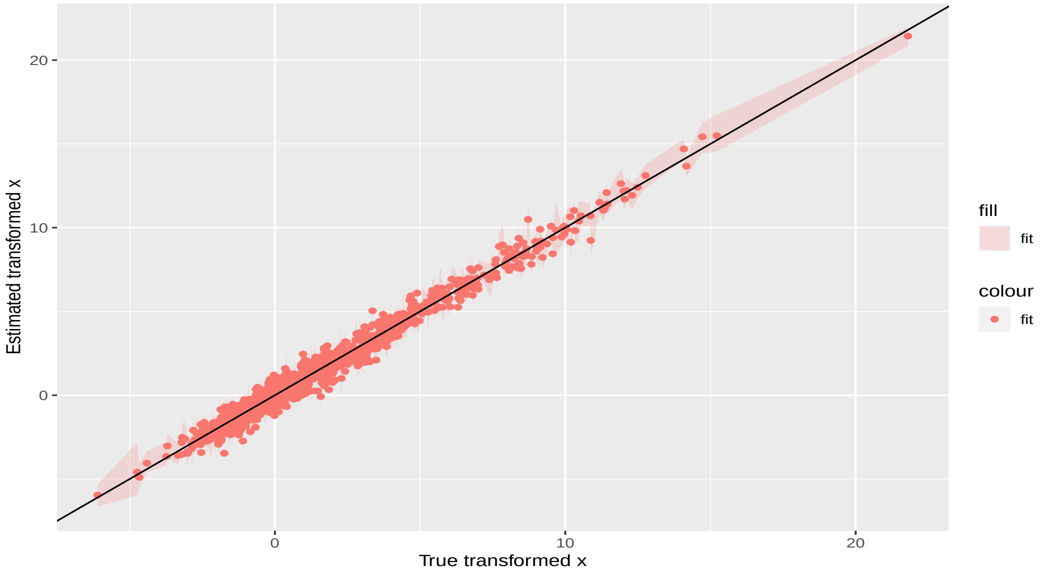
$$
\begin{aligned}
\boldsymbol{\theta} &\sim p(\boldsymbol{\theta}) \\
(\boldsymbol{x} \mid \boldsymbol{\theta}) &\sim \mathsf{N}(\boldsymbol{\mu}_x, \boldsymbol{Q}_x^{-1}) \\
(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta}) &\sim p(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta}) \\
\widehat{p}(\boldsymbol{\theta}|\boldsymbol{y}) &\propto \left. \frac{p(\boldsymbol{\theta})p(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta})}{\widehat{p}(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})} \right|_{\boldsymbol{x}=\widehat{\boldsymbol{x}}(\boldsymbol{\theta})} \qquad \text{(Laplace approximation)} \\
\widehat{p}(\boldsymbol{x}|\boldsymbol{y}) &= \int \widehat{p}(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y}) p(\boldsymbol{\theta}|\boldsymbol{y}) \,\mathrm{d}\boldsymbol{\theta}
\end{aligned}
$$

The usual INLA-able models have $\mathsf{E}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta}) = g^{-1}[\eta(\boldsymbol{x})]$ for some link function $g(\cdot)$ and linear predictor $\eta(\boldsymbol{x})$.
The `inlabru` package allows nonlinear $\eta(\boldsymbol{x})$, calling INLA iteratively on linearisations $\overline{\eta}(\boldsymbol{x}) = \eta(\boldsymbol{x}^*) + \nabla_x \eta(\boldsymbol{x})\big|_{\boldsymbol{x}=\boldsymbol{x}^*}^{\top} (\boldsymbol{x} - \boldsymbol{x}^*)$
Let's try this on the POQ transformation model with tail power parameters 0 (a skew-logistic model) and the others coded as transformed Gaussians.


20
10
0
Transformed x plus noise
−4
−2
0
2
True latent x



Estimated latent x
1
0
−1
−4
−2
0
2
True latent x
fill
fit
colour
fit


Estimated transformed x
20
10
0
0
10
20
True transformed x
fill
fit
colour
fit

What went wrong? Joint mode of (x, σ) is problematic! Toy example:

Transform $\mathsf{N}(0,1)$ variables to $\mathsf{N}(0,\sigma^2)$ variables, and observe with noise:

$$
\begin{aligned}
x_i &\sim \mathsf{N}(0,1), \quad i = 1,\dots,n \\
\sigma &\sim \mathsf{Exp}(\lambda) \\
(y_i|x_i,\sigma) &\sim \mathsf{N}(\sigma x_i, \sigma_\epsilon^2), \quad \text{known } \sigma_\epsilon > 0.
\end{aligned}
$$

Posterior log-density:

$$
\log p(\{x_i\},\sigma|\{y_i\}) = \text{const} - \frac{1}{2}\sum_i x_i^2 - \lambda\sigma - \frac{1}{2\sigma_\epsilon^2}\sum_i (y_i - \sigma x_i)^2
$$

Conditional modes $\widehat{x}_i = \frac{\sigma y_i}{\sigma^2+\sigma_\epsilon^2}$ gives profile log-posterior density

$$
\log p(\{\widehat{x}_i(\sigma)\},\sigma|\{y_i\}) = \text{const} - \lambda\sigma - \frac{1}{2(\sigma^2+\sigma_\epsilon^2)}\sum_i y_i^2 \approx \text{const} - \lambda\sigma - \frac{n(\sigma_{\text{true}}^2+\sigma_\epsilon^2)}{2(\sigma^2+\sigma_\epsilon^2)}
$$

When $\sigma_\epsilon \approx 0$, the maximum is at $\widehat{\sigma} \approx \left(\frac{n\sigma_{\text{true}}^2}{\lambda}\right)^{1/3}$
Note: Mode of marginal $p(\sigma|\{y_i\})$ is $\approx \sigma_{\text{true}}$.

INLA/inlabru integration of multi-parameter observation models

- Poisson point process likelihood

$$
\begin{aligned}
l(\boldsymbol{\theta}; y) &= -\int_\Omega \lambda(x|\boldsymbol{\theta})\,\mathrm{d}x + \sum_i \log[\lambda(y_i|\boldsymbol{\theta})] \\
&= -\int_\Omega \exp[\eta(x|\boldsymbol{\theta})]\,\mathrm{d}x + \sum_i \eta(y_i|\boldsymbol{\theta})
\end{aligned}
$$

- Can be linearised as $\eta(y|\boldsymbol{\theta}) = \log[\lambda(y)|\boldsymbol{\theta}] \approx \overline{\eta}(y|\boldsymbol{\theta}) = \eta(y|\boldsymbol{\theta}^*) + \left.\frac{\mathrm{d}\eta(y|\boldsymbol{\theta})}{\mathrm{d}\boldsymbol{\theta}}\right|_{\theta^*}^\top (\boldsymbol{\theta} - \boldsymbol{\theta}^*)$
- Implemented in inlabru via numerical integration schemes

$$
\widehat{l}(\boldsymbol{\theta}; y) = -\sum_j w_j \exp[\overline{\eta}(x_j|\boldsymbol{\theta})] + \sum_i \overline{\eta}(y_i|\boldsymbol{\theta})
$$

Attempting the "Poisson trick"

Given a density model $f(y|\boldsymbol{\theta})$ and $\phi \sim \text{Unif}(\mathcal{R})$, the "Poisson trick" reinterprets the model as a point process likelihood

$$
\begin{aligned}
l(\boldsymbol{\theta}; y) &= -\int_{\Omega} f(x|\boldsymbol{\theta}) e^{\phi} \,\mathrm{d}x + \log[f(y|\boldsymbol{\theta})] + \phi \\
&= -\sum_j \int_{\Omega_j} \exp\{\log[f(x|\boldsymbol{\theta})] + \phi\} \,\mathrm{d}x + \log[f(y|\boldsymbol{\theta})] + \phi
\end{aligned}
$$

where the $\{\Omega_j\}$ splits the domain into sub-intervals. The resulting posterior distribution for $\boldsymbol{\theta}$ is the same as for the original problem.
Rewrite:

$$
l(\boldsymbol{\theta}; y) = -\sum_j \exp\left[\eta_j(\boldsymbol{\theta}, \phi)\right] + \eta_y(\boldsymbol{\theta}, \phi)
$$

Problem: the $\eta_{\cdot}(\cdot)$ functions aren't monotonic in $\boldsymbol{\theta}$, so the iterated linearised $\overline{\eta}_{\cdot}(\cdot)$ version can be a poor approximation; can capture the correct posterior mode, but not the posterior variance

Likelihood contribution construction proof of concept

Example: Let $y \sim \mathsf{N}(\mu, \exp(2\theta))$, with likelihood contribution

$$
\begin{aligned}
l(\mu,\theta) = \log f(y|\mu,\theta) &= C_0 - \theta - \frac{1}{2}(y-\mu)^2 \exp(-2\theta) \\
&\approx C_0 - \theta - \frac{1}{2}(e^{a_0+a_1(\mu-\mu^*)} + e^{b_0+b_1(\mu-\mu^*)} - C_1)e^{-2\theta^*-2(\theta-\theta^*)} = \widehat{l}(\mu,\theta)
\end{aligned}
$$

where a_0, a_1, b_0, b_1, and C_1 are chosen so that $\widehat{l}(\mu,\theta)$ matches the first and second order derivatives at (μ^*,θ^*), and e.g. the modes match.
In the numerical Poisson point process likelihood construction, take

$$
\begin{aligned}
\eta_y(\mu,\theta) &= -\theta \\
\eta_1(\mu,\theta) &= -\log(2) + a_0 + a_1(\mu-\mu^*) - 2\theta \\
\eta_2(\mu,\theta) &= -\log(2) + b_0 + b_1(\mu-\mu^*) - 2\theta \\
\eta_3(\mu,\theta) &= -\log(2) + \log(C_1) - 2\theta
\end{aligned}
$$

Note however that $\exp(\eta_3)$ needs a *positive* sign: $\widehat{l}(\mu,\theta) = -e^{\eta_1} - e^{\eta_2} + e^{\eta_3} + \eta_y$
In this example, $\widehat{l}(\mu,\theta)$ can be chosen arbitrarily close to $l(\mu,\theta)$.

Summary

- Flexible parametric quantile models
- Transformed Gaussian processes/fields useful models, hard to estimate jointly, but step-wise or iterated approaches are feasible
- Likelihood construction or approximation for INLA possible, allowing flexible spatially and spatio-temporally varying observation models

Related reading:

- Vandeskog, S. M., Thorarinsdottir, T. L., Steinsland, I., Lindgren, F. (2022). Quantile based modeling of diurnal temperature range with the five-parameter lambda distribution. Environmetrics, 33(4), e2719. https://doi.org/10.1002/env.2719
- Fabian E. Bachl, Finn Lindgren, David L. Borchers, and Janine B. Illian (2019) *inlabru: an R package for Bayesian spatial modelling from ecological survey data*, Methods in Ecology and Evolution, 10(6):760–766. https://doi.org/10.1111/2041-210X.13168
- The INLA package; https://www.r-inla.org
- The inlabru package; https://inlabru-org.github.io/inlabru/