Taking LGMs beyond GLMs in climate and ecology

Finn Lindgren
finn.lindgren@ed.ac.uk

THE UNIVERSITY *of* EDINBURGH

School of Mathematics

10th November 2016

Out there, there's a world outside of Yonkers

THE UNIVERSITY *of* EDINBURGH

The strict GLM/GAM/GLMM/etc subset of Latent Gaussian Models is limited in comparison with hierarchical models with non-linear links between nodes of latent Gaussian random processes and fields.

Prior and conditional posterior approximation at the core of INLA:

$$
\begin{aligned}
\log p(\boldsymbol{x}|\boldsymbol{\theta}) &= C_x - \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{Q}_x \boldsymbol{x} \\
\log p(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y}) &= \log p(\boldsymbol{x}|\boldsymbol{\theta}) + \log p(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{x}) \\
&\approx C_{x|y} - \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_{x|y})^\top \boldsymbol{Q}_{x|y}(\boldsymbol{x}-\boldsymbol{\mu}_{x|y}) \\
\boldsymbol{Q}_x \boldsymbol{\mu}_{x|y} &= \left[\nabla_{\boldsymbol{x}} \log p(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{x})\right]_{\boldsymbol{x}=\boldsymbol{\mu}_{x|y}} \\
\boldsymbol{Q}_{x|y} &= \boldsymbol{Q}_x - \left[\nabla_{\boldsymbol{x}}\nabla_{\boldsymbol{x}}^\top \log p(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{x})\right]_{\boldsymbol{x}=\boldsymbol{\mu}_{x|y}}
\end{aligned}
$$

How far can one extend this Gaussian approximation technique?

Put on your Sunday clothes there's lots of world out there

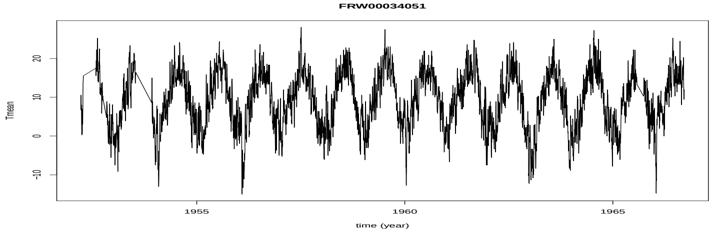
THE UNIVERSITY *of* EDINBURGH

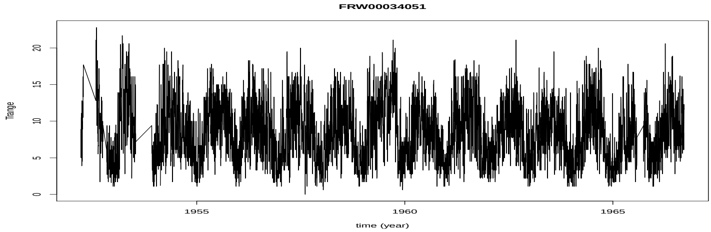
Some small steps in this direction:

- Products of spectrally separated processes; climate and weather
- Partially observed LGCPs; imperfect point detection
- Mark-dependent detection probabilities; 500 dolphins in a group are more visible than 5

Observed data

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

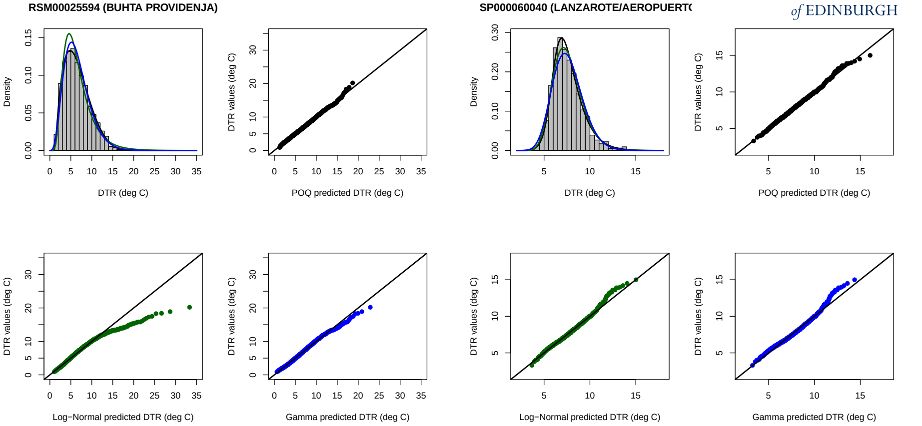




Reasonable models:

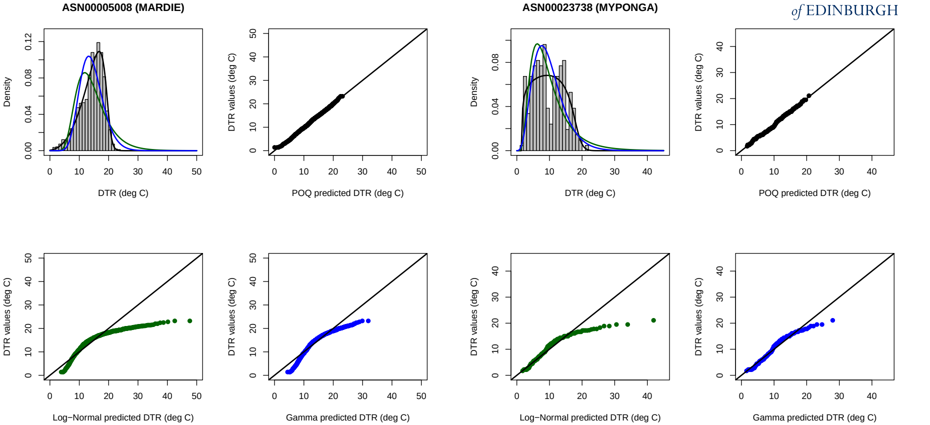
$T_m(t) = T_m^c(t) + T_m^a(t)$ and $T_r(t) = \mathrm{e}^{T_r^c(t)} h_t[T_r^a(t)]$

Diurnal range distributions



For these stations, POQ does a slightly better job than a Gamma distribution.

Diurnal range distributions



For these stations only POQ comes close to representing the distributions. Note: Some of the mixture-like distribution shapes may be an effect of unmodeled station inhomogeneities.

Products of transformed processes

THE UNIVERSITY *of* EDINBURGH

Assume that $\boldsymbol{u}$ is a large scale process and $\boldsymbol{v}$ is a small scale process, so that they are statistically identifiable from observations of the form

$$y_i = h_u(u_i) \cdot h_v(v_i) + \epsilon_i, \quad h_u \text{ and } h_v \text{ non-linear transformations.}$$

Write $\boldsymbol{h}_u$, $\boldsymbol{h}_u'$, $\boldsymbol{h}_u''$ for the vectors of transformed values and derivatives of h_u at the u_i values, and similarly for $\boldsymbol{v}$. Then

$$\begin{aligned}
C - \log p(\boldsymbol{y} \mid \boldsymbol{u}, \boldsymbol{v}) &= \frac{1}{2}(\boldsymbol{y} - \boldsymbol{h}_u \odot \boldsymbol{h}_v)^\top \boldsymbol{Q}_\epsilon (\boldsymbol{y} - \boldsymbol{h}_u \odot \boldsymbol{h}_v) \\
-\frac{\partial}{\partial \boldsymbol{v}} \log p(\boldsymbol{y} \mid \boldsymbol{u}, \boldsymbol{v}) &= -\operatorname{diag}(\boldsymbol{h}_u \odot \boldsymbol{h}_v') \boldsymbol{Q}_\epsilon (\boldsymbol{y} - \boldsymbol{h}_u \odot \boldsymbol{h}_v) \\
-\frac{\partial^2}{\partial \boldsymbol{v}^2} \log p(\boldsymbol{y} \mid \boldsymbol{u}, \boldsymbol{v}) &= \operatorname{diag}(\boldsymbol{h}_u \odot \boldsymbol{h}_v') \boldsymbol{Q}_\epsilon \operatorname{diag}(\boldsymbol{h}_u \odot \boldsymbol{h}_v') \\
&\quad - \operatorname{diag}(\operatorname{diag}(\boldsymbol{h}_u \odot \boldsymbol{h}_v'') \boldsymbol{Q}_\epsilon (\boldsymbol{y} - \boldsymbol{h}_u \odot \boldsymbol{h}_v))
\end{aligned}$$

and similarly for $\frac{\partial}{\partial \boldsymbol{u}}$, $\frac{\partial^2}{\partial \boldsymbol{u} \partial \boldsymbol{v}}$, and $\frac{\partial^2}{\partial \boldsymbol{u}^2}$. The problematic term in the Hessian involving $\boldsymbol{y}$ disappears in Fisher scoring: $\mathsf{E}_{\boldsymbol{y}|\boldsymbol{u},\boldsymbol{v}} \left(-\nabla^2_{(\boldsymbol{u},\boldsymbol{v})} \ln p(\boldsymbol{y} \mid \boldsymbol{u}, \boldsymbol{v}) \right)$ is positive definite.

LGCPs in INLA

THE UNIVERSITY *of* EDINBURGH

An inhomogeneous point process $Y = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N(\Omega)\}$ on a space Ω with intensity $\lambda(\mathbf{s})$, $\mathbf{s} \in \Omega$ is defined so that for each region $A \subset \Omega$, the number of points is $N_Y(A) \sim \mathsf{Po}(\int_A \lambda(\mathbf{s})\,\mathrm{d}\mathbf{s})$.

A Log-Gaussian Cox Process has a log-linear latent Gaussian (spatial) model for $\lambda(\mathbf{s}) = \exp(\eta(\mathbf{s}))$.

The conditional likelihood is

$$\begin{aligned}\log p(\boldsymbol{Y} | \eta(\cdot), \boldsymbol{\theta}) &= -\int_\Omega \lambda(\mathbf{s})\,\mathrm{d}\mathbf{s} + \sum_{i=1}^{N_Y(\Omega)} \log \lambda(\boldsymbol{y}_i) \\ &\approx -\sum_{j=1}^{J} w_j \lambda(\mathbf{s}_j) + \sum_{i=1}^{N_Y(\Omega)} \log \lambda(\boldsymbol{y}_i)\end{aligned}$$

where $\mathbf{s}_1, \ldots, \mathbf{s}_J$ are numerical integration points with corresponding integration weights w_j.

Partially observed LGCPs

THE UNIVERSITY *of* EDINBURGH

In *line transect surveys*, the probability $g(z(\mathbf{s}))$ of detecting a *point* (a group of animals) at location $\mathbf{s}$ depends on the distance to the observer line, $z(\mathbf{s})$.

The region with non-zero detection probability, $\widetilde{\Omega}$, may be much smaller than Ω.

This results in a *thinned* point process with intensity $\lambda(\mathbf{s})g(z(\mathbf{s}))$ on $\widetilde{\Omega}$

If $g(\cdot)$ is a *half-normal detection function*,

$$\log g(z;\beta) = -\beta z^2/2,$$

the thinned intensity is log-linear in both $\eta(\cdot)$ and β, so the combined model can use a variant of the numerical implementation used for the completely observed model.

What about the *hazard rate model*,

$$g(z;\gamma,\sigma) = 1 - \exp\left[-\left(\frac{z}{\sigma}\right)^{-\gamma}\right],$$

?

Linearisation

THE UNIVERSITY *of* EDINBURGH

For some current MAP estimates $\widehat{\beta}$ and $\log\widehat{\sigma}$, a first order Taylor series expansion gives

$$
\begin{aligned}
\log g(z;\beta,\sigma) \approx{}& \log g(z;\widehat{\beta},\widehat{\sigma}) \\
&+ (\beta-\widehat{\beta})\left[\frac{\partial}{\partial\beta}\log g(z;\beta,\sigma)\right]_{(\beta,\sigma)=(\widehat{\beta},\widehat{\sigma})} \\
&+ (\log\sigma-\log\widehat{\sigma})\left[\frac{\partial}{\partial\log\sigma}\log g(z;\beta,\sigma)\right]_{(\beta,\sigma)=(\widehat{\beta},\widehat{\sigma})} \\
\approx{}& \widetilde{g}_{\widehat{\beta},\widehat{\sigma}}(z) + (\beta-\widehat{\beta})\widetilde{g}^{\beta}_{\widehat{\beta},\widehat{\sigma}}(z) + (\log\sigma-\log\widehat{\sigma})\widetilde{g}^{\log\sigma}_{\widehat{\beta},\widehat{\sigma}}(z) \\
={}& \log\widetilde{g}_{\widehat{\beta},\widehat{\sigma}}(z(s);\beta,\sigma)
\end{aligned}
$$

This linearisation allows us to treat β and $\log\sigma$ as *fixed effects*. Generalising to spatially varying parameter fields is also permitted, e.g. with a Gaussian process prior on $\log\sigma(s)$

Estimation is carried out by iterated calls to `inla()` with the LGCP log-linear intensity model $\lambda(\mathbf{s})\widetilde{g}_{\widehat{\beta},\widehat{\sigma}}(z(\mathbf{s});\beta,\sigma)$

Mark dependent detection probability

The probability of detecting a group of animals at a large distance is larger for large groups than for small groups. Failure to model that leads to bias.

We can design a group size dependent detection function, e.g. $g(z, m) = 1 - \exp\left[-\left(\frac{z}{\sigma(m)}\right)^{-\beta}\right]$, with $\log \sigma(m) = \alpha + \beta m$,

but the log-groupsize m is only available where we have *detected* a group of animals. We need to model what the group size *could potentially be* at all locations in the studied domain.

A simple model is the continuous log-groupsize model $(m|\mu(\cdot), \mathbf{s}) \sim \mathsf{N}(\mu(\mathbf{s}), 1/\tau)$

The joint point process for detected points and log-groupsizes on $(\mathbf{s}, m) \in \widetilde{\Omega} \times \mathbb{R}$ has intensity $\lambda(\mathbf{s}) p(m|\mu(\cdot), \mathbf{s}) g(z(\mathbf{s}), m)$

Linearised likelihood

THE UNIVERSITY *of* EDINBURGH

The conditional likelihood is

$$\log p(\boldsymbol{Y}, \boldsymbol{M} | \eta(\cdot), \mu(\cdot), \boldsymbol{\theta}, \tau) = -\int_{\widetilde{\Omega}} \int_{\mathbb{R}} \lambda(\mathbf{s}) p(m | \mu(\mathbf{s}), \tau) g(z(\mathbf{s}); \boldsymbol{\theta}) \, \mathrm{d}m \, \mathrm{d}\mathbf{s}$$

$$+ \sum_{i=1}^{N_Y(\Omega)} \log \lambda(\boldsymbol{y}_i) + \sum_{i=1}^{N_Y(\Omega)} \log p(m_i | \mu(\boldsymbol{y}_i), \tau) + \sum_{i=1}^{N_Y(\Omega)} \log g(z(\boldsymbol{y}_i); \boldsymbol{\theta})$$

$$\approx -\sum_{j=1}^{J} w_j \lambda(\mathbf{s}_j) \widetilde{p}(m_j | \mu(\mathbf{s}_j), \tau) \widetilde{g}(z(\mathbf{s}_j); \boldsymbol{\theta})$$

$$+ \sum_{i=1}^{N_Y(\Omega)} \log \lambda(\boldsymbol{y}_i) + \sum_{i=1}^{N_Y(\Omega)} \log \widetilde{p}(m_i | \mu(\boldsymbol{y}_i), \tau) + \sum_{i=1}^{N_Y(\Omega)} \log \widetilde{g}(z(\boldsymbol{y}_i); \boldsymbol{\theta})$$

where $(\mathbf{s}_1, m_1), \ldots, (\mathbf{s}_J, m_J)$ are numerical integration points with corresponding integration weights w_j, and $\log \widetilde{p}(m | \mu(\cdot), \tau)$ and $\log \widetilde{g}(z(\mathbf{s}); \boldsymbol{\theta})$ are 1st order Tayor approximations at some $\widehat{\mu}(\cdot)$, $\widehat{\tau}$, and $\widehat{\boldsymbol{\theta}}$.

General principle

THE UNIVERSITY *of* EDINBURGH

We can gain some intuition why the linearisation works by reinterpreting an ordinary likelihood $\prod_{i=1}^{n} p(y_i|x)$ as a point process intensity. Define

$$\gamma_{\widehat{x}}(y) = \left[\frac{\partial}{\partial x}\log p(y|x)\right]_{x=\widehat{x}}$$

and take a 1st order Taylor approximation of $\log p(\boldsymbol{y}|x)$:

$$\begin{aligned}
\log p(\boldsymbol{y}|x) &= n - n\int p(y|x)\,\mathrm{d}y + \sum_{i=1}^{n}\log p(y_i|x) \\
&\approx n - n\int p(y|\widehat{x})\mathrm{e}^{(x-\widehat{x})\gamma_{\widehat{x}}(y)}\,\mathrm{d}y \\
&\quad + \sum_{i=1}^{n}\log p(y_i|\widehat{x}) + (x-\widehat{x})\sum_{i=1}^{n}\gamma_{\widehat{x}}(y_i)
\end{aligned}$$

Trivia: The integral term is the mgf for $\gamma_{\widehat{x}}(y)$ evaluated at $x-\widehat{x}$.

General principle

The derivatives can be evaluated at $x = \widehat{x}$:

$$\frac{\partial}{\partial x} \log p(\boldsymbol{y}|x) = -n \int \gamma_{\widehat{x}}(y) p(y|\widehat{x}) \mathrm{e}^{(x-\widehat{x})\gamma_{\widehat{x}}(y)} \,\mathrm{d}y + \sum_{i=1}^{n} \gamma_{\widehat{x}}(y_i)$$
$$= -n\mathsf{E}\left[\gamma_{\widehat{x}}(y) \mid y \sim p(y|\widehat{x})\right] + \sum_{i=1}^{n} \gamma_{\widehat{x}}(y_i) = \sum_{i=1}^{n} \gamma_{\widehat{x}}(y_i)$$

which is the same derivative as for $\log p(\boldsymbol{y}|x)$ at $\widehat{x}$.

$$\frac{\partial^2}{\partial x^2} \log p(\boldsymbol{y}|x) = -n \int \gamma_{\widehat{x}}(y)^2 p(y|\widehat{x}) \mathrm{e}^{(x-\widehat{x})\gamma_{\widehat{x}}(y)} \,\mathrm{d}y$$
$$= -n\mathsf{E}\left[\gamma_{\widehat{x}}(y)^2 \,\middle|\, y \sim p(y|\widehat{x})\right] = -\mathcal{I}(\widehat{x}|\boldsymbol{y})$$

where $\mathcal{I}(\widehat{x})$ is the Fisher information for x evaluated at $\widehat{x}$.

The computational loss in using this approximation is the numerical integration over y.

Bonus preview: Interactive mesh builder

THE UNIVERSITY *of* EDINBURGH

`shiny` app for experimenting with and assessing meshes for spde models.
SD on the mesh, continuous domain SD, and their ratio: