Bayesian hierarchical modeling of extreme hourly precipitation in Norway

Alex Lenkoski
Norsk Regnesentral/Norwegian Computing Center

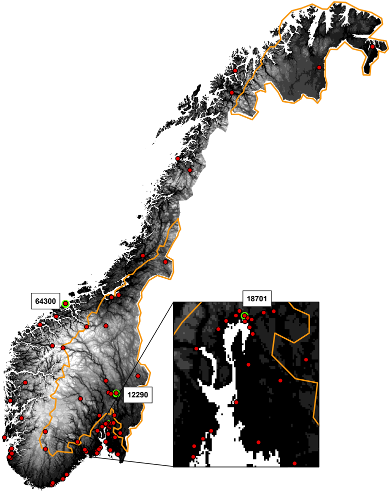
Joint work with Thordis Thorarinsdottir, Anita Dyrrdal, Frode Stordal

August 28, 2014

Return Model of Extreme Precipitation in Norway

Data Collection

Annual maximum 24-hour precipitation at 69 sites in Norway collected between 1965-2012



Hierarchical Spatial GEV Models

Let Y_{ts} be the annual max 24-hour precip in year t at location s. We assume

$$Y_{ts} \sim GEV(\mu_s, \kappa_s, \xi_s)$$

where

$$pr(Y_{ts}|\mu_s, \xi_s, \kappa_s) = \kappa_s h(Y_{ts})^{-(\xi_s+1)/\xi_s} \exp\left\{-h(Y_{ts})^{-\xi_s^{-1}}\right\}$$

with

$$h(Y_{ts}) = 1 + \xi_s \kappa_s (Y_{ts} - \mu_{ts}).$$

Our goal is to thus model (μ_s, κ_s, ξ_s) throughout Norway, given the observations at the 69 locations and calculate the $1/p$ return level

$$z_p^s = \mu_s - \frac{1}{\kappa_s \xi_s}\left\{1 - [-\log(1-p)]^{-\xi_s}\right\}$$

Hierarchical Spatial GEV Model

In order to impute values beyond our observations we specify a hierarchial model

$$Y_{ts} \sim GEV(\mu_s, \kappa_s, \xi_s)$$

such that

$$\mu_s = \boldsymbol{\theta}'_\mu \mathbf{X}_s + \tau_s^\mu$$
$$\tau_s^\mu \sim \mathcal{GP}(\alpha_\mu, \lambda_\mu)$$

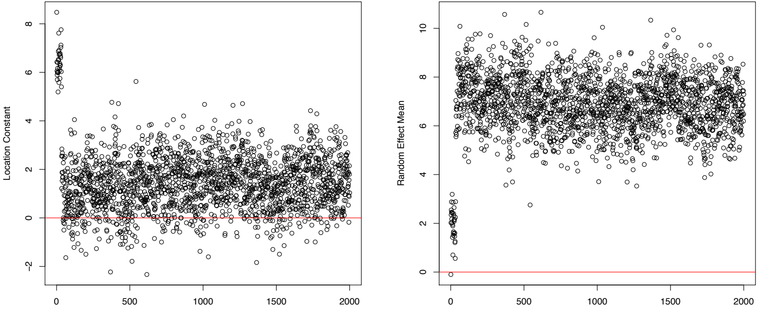
where

$$cov(\tau_s, \tau_{s'}) = \alpha_\mu^{-1} \exp\left(-d_{ss'}/\lambda_\mu\right)$$
$$\alpha_\mu \sim \Gamma(a_\alpha^\mu, b_\alpha^\mu)$$
$$\lambda_\mu \sim \Gamma(a_\lambda^\mu, b_\lambda^\mu)$$

And similar models for κ_s, ξ_s.

How to turn Scientists off Bayes

The model above is outlined in Davison et al. (Stat Sci, 2012) with an implementation in an existing R package. Our collaborators contacted us after getting results similar to those below

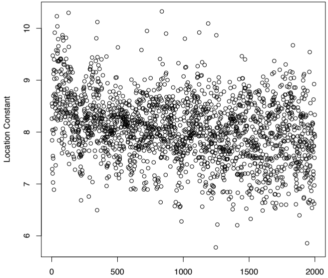


(a) Constant Term (b) RE Mean

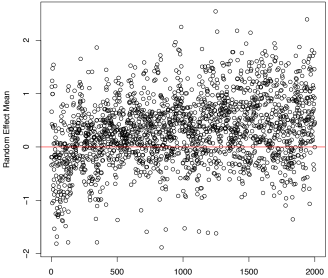
The fix was easy–was it Bayesian?

We solve this by setting

$$E\{\boldsymbol{\theta}_\mu\} = (8, 0, \ldots, 0)$$



(c) Constant Term



(d) RE Mean

Have I cheated? Certainly, our collaborators didn't care.

BMA and Conditional Bayes Factors

Now let M^{μ} be a model restricting certain elements of $\boldsymbol{\theta}_{\mu}$ to zero (and M^{κ}, M^{ξ} be similar). Suppose we're in the course of an MCMC and set

$$\boldsymbol{\Upsilon}^{\mu} = \mathbf{X}\boldsymbol{\theta}_{\mu} + \boldsymbol{\tau}^{\mu}.$$

We then update M^{μ} and $\boldsymbol{\theta}_{\mu}$ as a block, by noting

$$pr(M_{\mu}|\boldsymbol{\Upsilon}_{\mu}, \mathbf{X}^{\mu}_{\mathcal{S}_o}) \propto |\boldsymbol{\Xi}_{\mu}|^{-1/2} \exp\left(\frac{1}{2}\hat{\boldsymbol{\theta}}'_{\mu}\boldsymbol{\Xi}_{\mu}\hat{\boldsymbol{\theta}}_{\mu}\right) pr(M_{\mu})$$
$$\boldsymbol{\Xi}_{\mu} = \mathbf{X}'\mathcal{K}^{-1}_{\alpha_{\mu},\lambda_{\mu}}\mathbf{X} + \boldsymbol{\Xi}_0$$
$$\hat{\boldsymbol{\theta}}_{\mu} = \boldsymbol{\Xi}^{-1}_{\mu}(\mathbf{X}'\mathcal{K}^{-1}_{\alpha_{\mu},\lambda_{\mu}}\boldsymbol{\Upsilon}_{\mu} + \boldsymbol{\Xi}_0\boldsymbol{\theta}_0)$$

We can quickly embed BMA inside hierarchical models using this conditional Bayes factor for transitions.

Further Developments – Better Metropolis Proposals

The original software required the specification of 9 M-H tuning parameters–a daunting task for anyone. Consider updating, e.g. τ_s^ξ and write

$$pr(\tau_s^\xi|\cdot) \propto \exp\{g(\tau_s^\xi)\}$$

We determine g' and g'' (a mess), set

$$b(\tau_s^\xi) = g'(\tau_s^\xi) - g''(\tau_s^\xi)\tau_s^\xi$$
$$c(\tau_s^\xi) = -g''(\tau_s^\xi)$$

and sample

$$(\tau_s^\xi)' \sim \mathcal{N}(b(\tau_s^\xi)/c(\tau_s^\xi), 1/c(\tau_s^\xi))$$

see, e.g. Rue and Held (2005).

Updating the Gaussian Process Parameters

Let $\mathbf{D}$ be the distance matrix, $\mathbf{E}(\lambda) = [\exp(-d_{ij}/\lambda)]$ and write

$$\begin{aligned}
\mathbf{F}(\lambda) &= \frac{\partial}{\partial\lambda}\mathbf{E}(\lambda) = \frac{1}{\lambda^2}\mathbf{D}\circ\mathbf{E}(\lambda) \\
\mathbf{G}(\lambda) &= \frac{\partial}{\partial\lambda}\mathbf{F}(\lambda) = -\frac{2}{\lambda^3}[\mathbf{D}\circ\mathbf{E}(\lambda)] + \frac{1}{\lambda^2}[\mathbf{D}\circ\mathbf{F}(\lambda)].
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial\lambda}\log|\mathbf{E}(\lambda)| &= tr\left\{\mathbf{E}^{-1}(\lambda)\mathbf{F}(\lambda)\right\} \\
\frac{\partial}{\partial\lambda}\boldsymbol{\tau}'\mathbf{E}(\lambda)^{-1}\boldsymbol{\tau} &= \boldsymbol{\tau}'\left(\mathbf{E}(\lambda)^{-1}\left[-\frac{\partial}{\partial\lambda}\mathbf{E}(\lambda)\right]\mathbf{E}(\lambda)^{-1}\right)\boldsymbol{\tau} \\
&= \boldsymbol{\tau}'\left(\mathbf{E}(\lambda)^{-1}[-\mathbf{F}(\lambda)]\mathbf{E}(\lambda)^{-1}\right)\boldsymbol{\tau}.
\end{aligned}$$

And so on (it gets tedious). This allows us to perform the same M-H as in the random effects case.

Spatial Interpolation

We run a Markov Chain which returns a collection of samples

$$\left\{\theta_\nu, \alpha_\nu, \lambda_\nu, \tau^\nu_{\mathcal{S}_o}\right\}^{[r]}_{\nu \in \{\mu,\kappa,\xi\}}$$

for $r = 1, \ldots, R$. Let $s \in \mathcal{S} \setminus \mathcal{S}_o$. We determine, e.g.,

$$\mu_s^{[r]} = (\boldsymbol{\theta}_\mu^{[r]})' \mathbf{X}_s + \tau_s^{[r]}$$

where $\tau_s^{[r]}$ is sampled from the $\mathcal{GP}$ model with $\alpha_\mu^{[r]}, \lambda_\mu^{[r]}$ as parameters conditional on the current random effects $(\tau^\mu_{\mathcal{S}_o})^{[r]}$. This yields a sample

$$\{(z_p^s)^{[1]}, \ldots, (z_p^s)^{[R]}\}$$

which approximates

$$pr(z_p^s | \mathbf{Y}_{\mathcal{S}_o}),$$

the posterior distribution of the pth return level at site s.

Some Results

Results from 200k reps after 20k burn-in (∼10 min run time).
Acceptance probabilities were great

	λ	Worst τ	Mean τ	Best τ
μ	0.852	0.850	0.951	1.00
κ	0.815	0.922	0.970	1.00
ξ	0.823	0.781	0.939	1.00

Some variable results (12 variables considered in total).

	μ		κ		ξ	
	Prob	ESS	Prob	ESS	Prob	ESS
Constant	1	28,155	1	165,263	1	152,291
Lat	0.47	8,009	0.15	44,564	0.13	102,567
Lon	0.42	12,302	0.16	96,424	0.12	38,652
Elevation	0.86	27,435	0.03	39,884	0.03	59,567
Dist from Sea	0.56	16,478	0.01	905,918	0.02	1,496,915

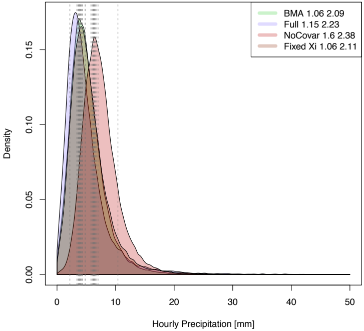
Out of sample predictive distributions

Using a leave-one-out cross validation exercise, compared our method (BMA) to the full model (Full), a method with only spatial information (NoCovar) and a method that set $\xi = .15$, (Fixed).

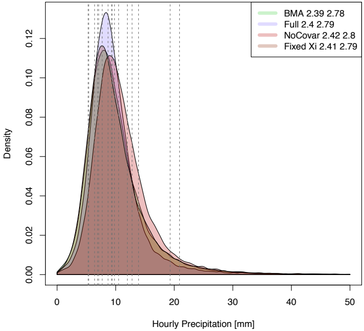
Table : Predictive Scores for the leave-one-out cross validation study

	CRPS	LS
BMA	2.520	2.823
Full	2.543	2.840
NoCovar	2.542	2.839
Fixed	2.525	2.826

Example Predictive Distributions



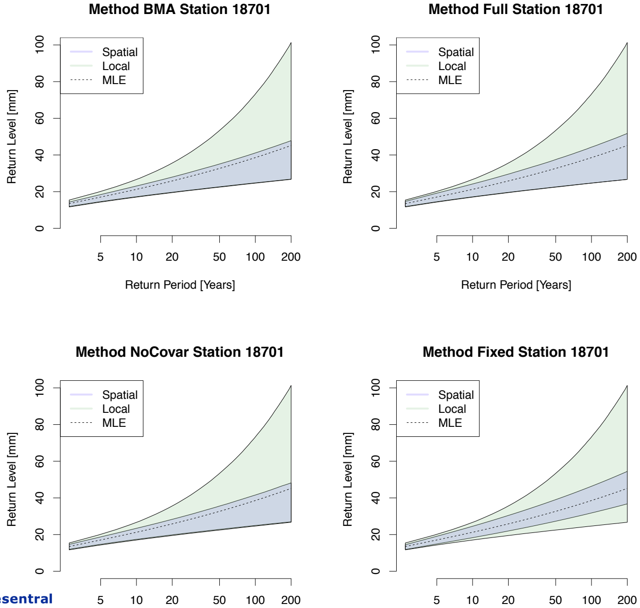
(e) Site 15720



(f) Site 40880

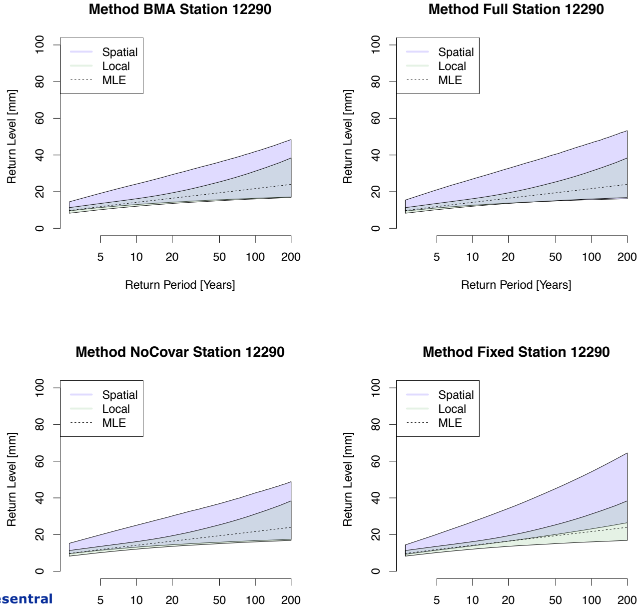
Norsk Regnesentral
NORWEGIAN COMPUTING CENTER

Return Levels for Oslo Fjord Station

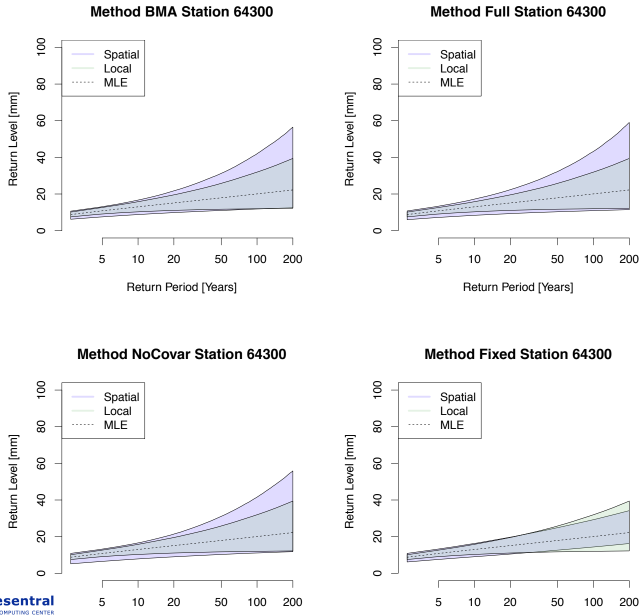


Norsk Regnesentral
NORWEGIAN COMPUTING CENTER

Return Levels for Mountain Station

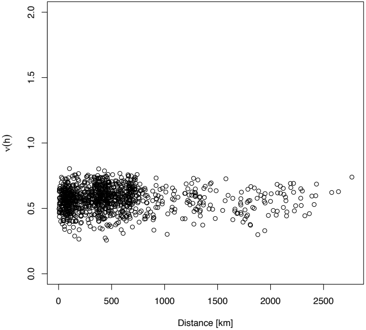


Return Levels for West Coast Station

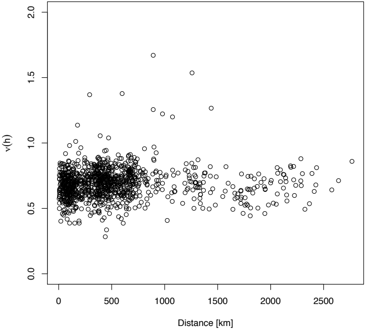


Madograms show little residual correlation

A Madogram (Cooley et al. 2006) indicates residual spatial dependence in models for extremes



(s) Empirical

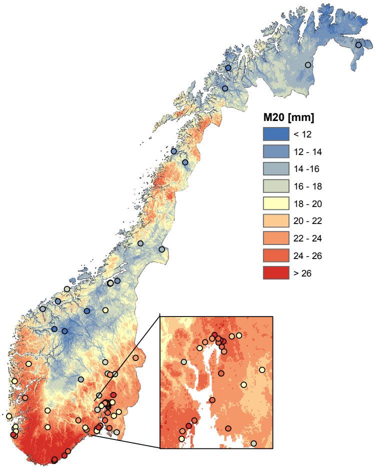


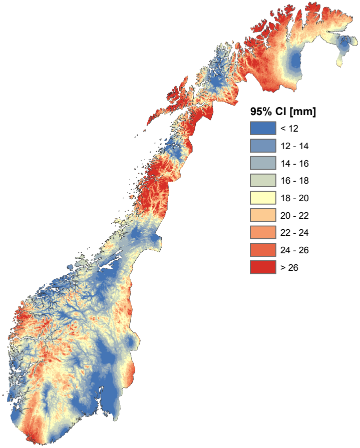
(t) MLE

Almost too little correlation?

Return Level Maps

We then use to determine the, say, 20-year return level





Discussion of Return Level Modeling

- A.V. Dyrrdal, A. Lenkoski, T. L. Thorarinsdottir, F. Stordal (2014). Bayesian hierarchical modeling of extreme hourly precipitation in Norway. *To appear, Environmetrics*
- When we don't fix the covariance of the GP we observe "identification" issues unless sensible priors are stipulated–how do characterize/resolve this in general?
- Nonstationarity and "semi-supervision": we have detailed precipitation information on a fine grid for Norway (both NWP output and observed). Could I use this as a "semi-supervised" method of constructing a non-stationary covariance function over Norway?
- A multiresolution Gaussian Process over regions of Norway?