Can a regional climate model reproduce observed extreme temperatures?

Peter F. Craigmile

The Ohio State University department of STATISTICS

http://www.stat.osu.edu/~pfc/

Workshop on the analysis of environmental extremes,

The Pennsylvania State University, University Park, PA.

24-25 Oct 2016.

Joint with Peter Guttorp at the University of Washington

and the Norwegian Computing Center.

P. F. Craigmile and P. Guttorp (2013). Can a regional climate model reproduce observed extreme temperatures? *Statistica*, **73**, 103-122.

My research is supported by the US National Science Foundation under grants NSF-DMS-1407604 and NSF-SES-1424481.

Investigating extremes in climate

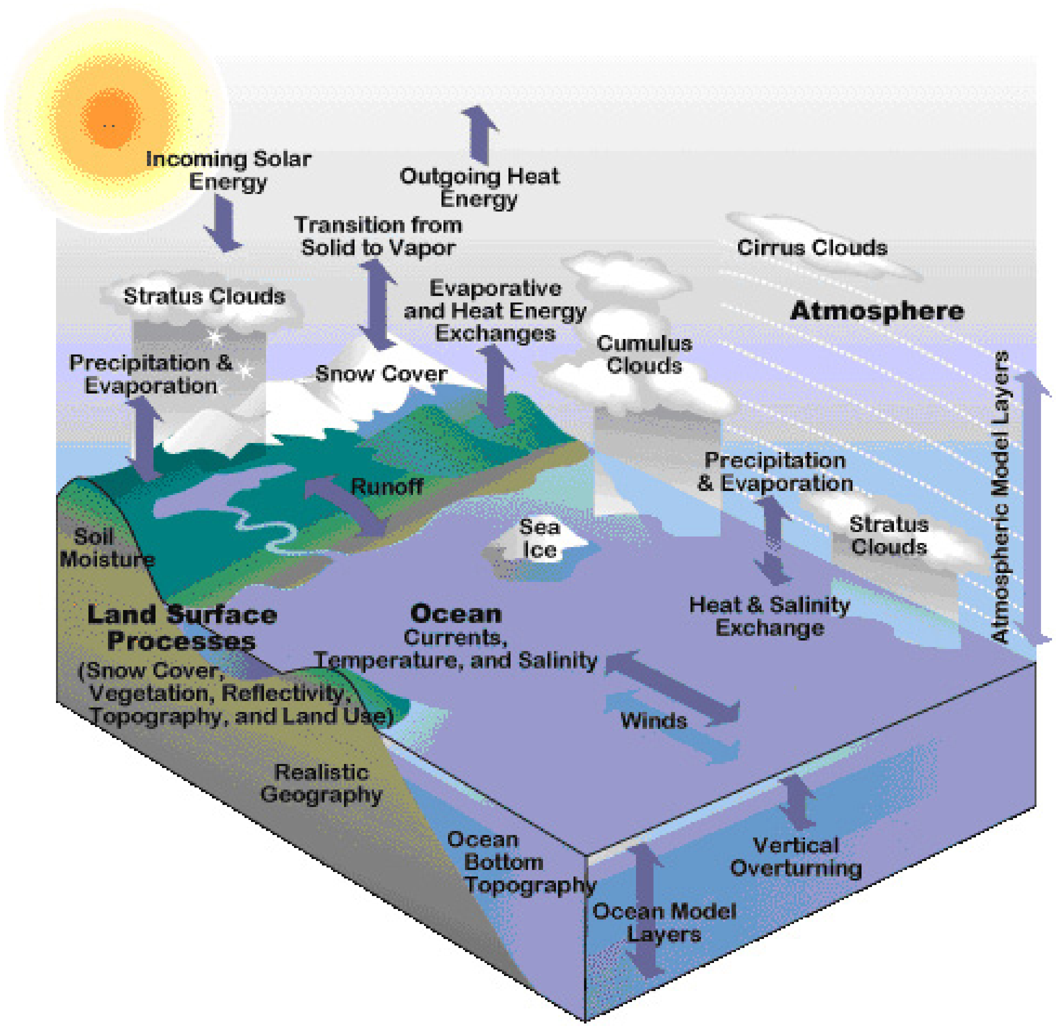
In order to adapt to a changing climate, policymakers need information about what to expect for the climate system, including **extremes**.

Local information about extremes:

- **observational measurements**
- **regional climate models (RCMs)**

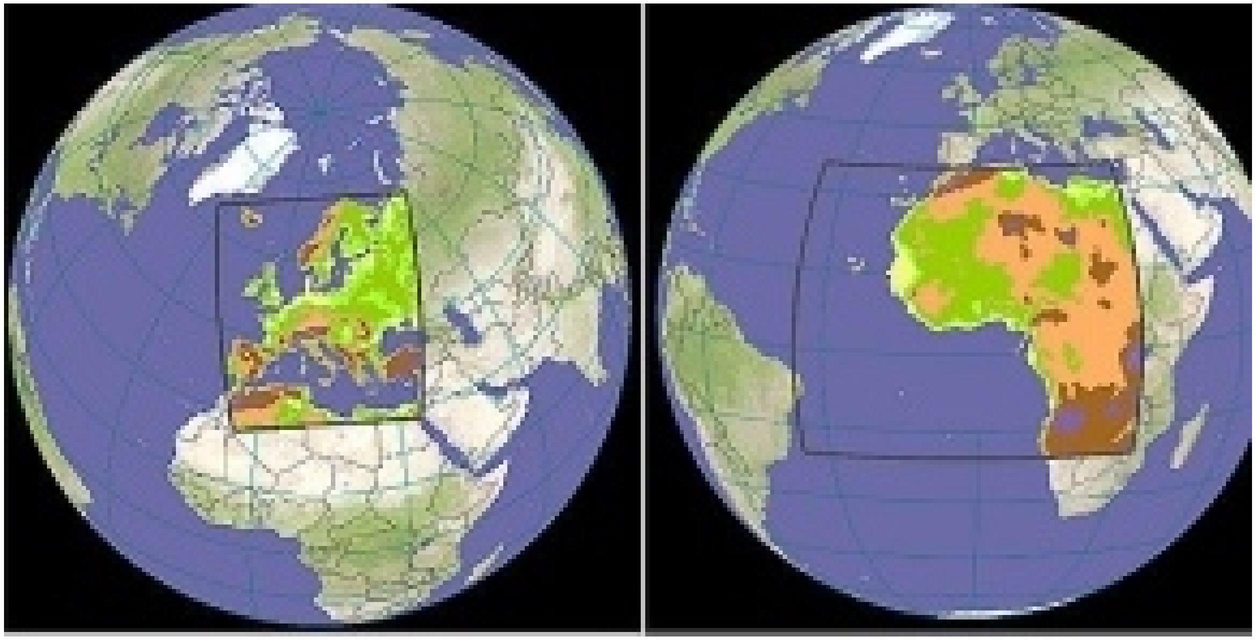
Q: How well do regional models **reproduce** observed **extremes** in climate?

Climate models



Source: http://www.windows2universe.org/earth/climate/cli_models2.html

Regional climate models

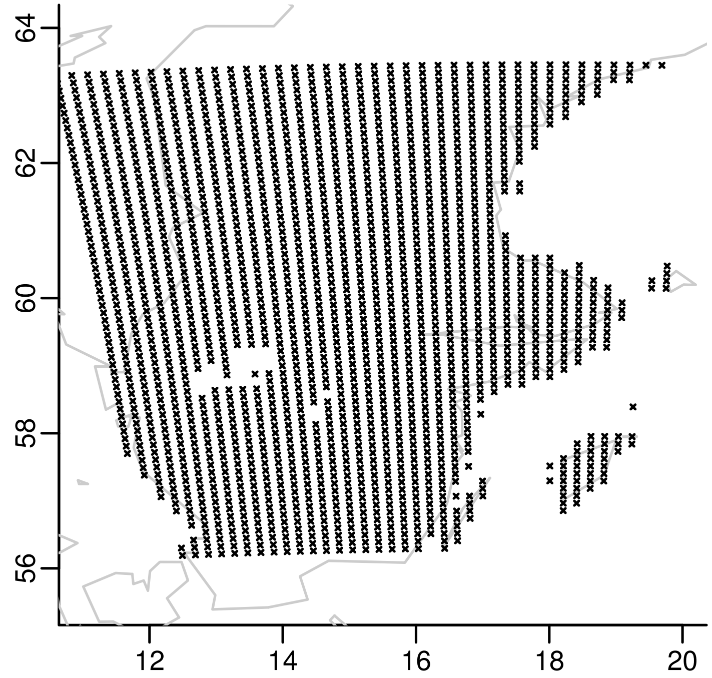


Source: http://www.smhi.se/en/research/research-news/combined-science-on-climate-models-1.14533

- RCMs are a downscaled global circulation/climate model (GCM).

  Mathematical model that describes, using partial differential equations, the temporal evolution of climate, oceans, atmosphere, ice, and land-use processes over a gridded spatial domain of interest.

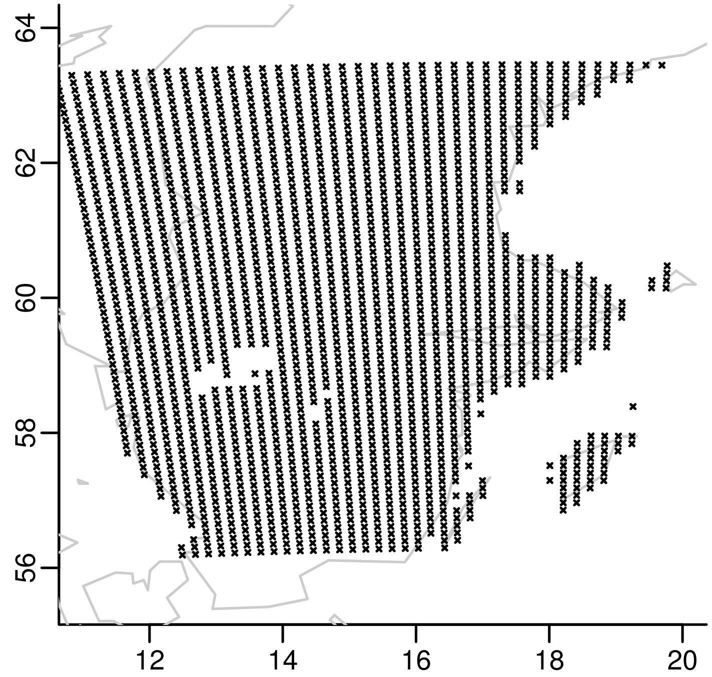
RCM output



A control run of the SMHI regional model RCA3 [Samuelsson et al., 2011]:

- Run using boundary conditions given by the ERA40 reanalysis [Uppala et al., 2005] in the earlier years, and the ERA-INTERIM reanalysis [Uppala et al., 2008] in the later years.

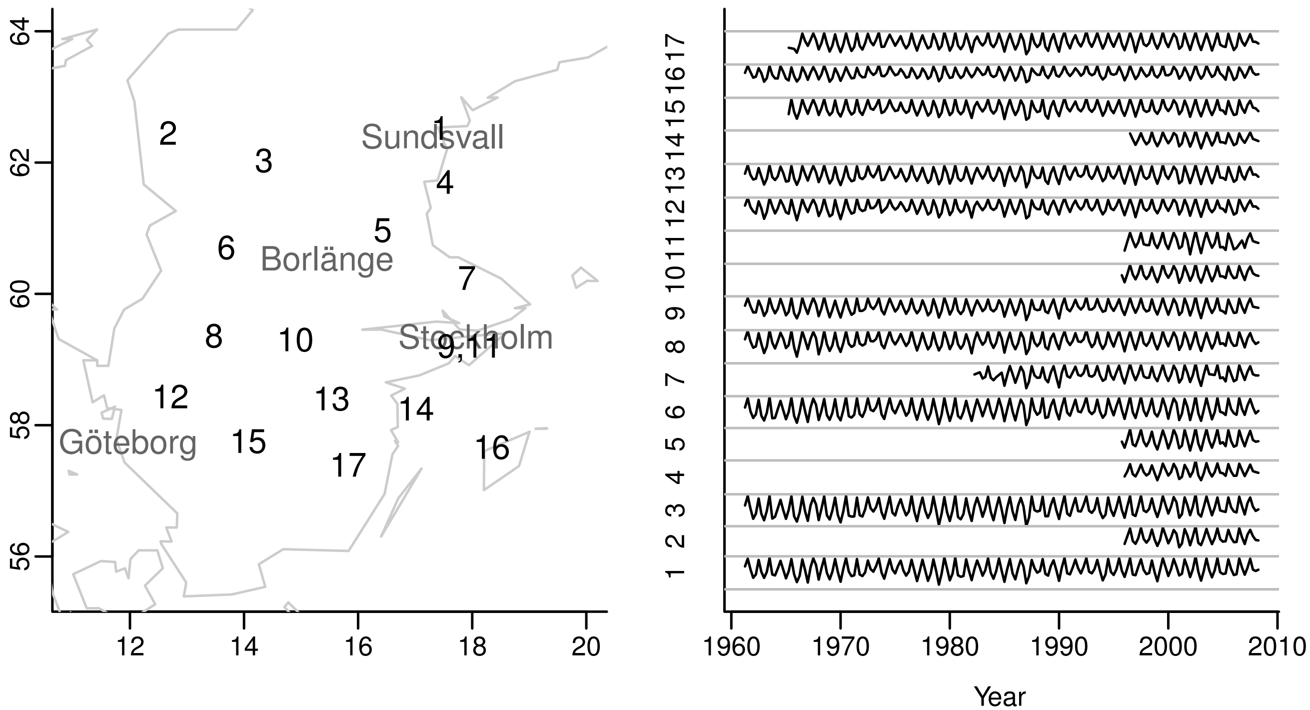
RCM output, continued



We restrict our analysis to the 2 meter temperature given for the **open land** and **snow land covers**.

Temporal resolution of 7.5 minutes, on a 12.5×12.5 km^2 grid – based on daily output values we calculate **seasonal minima** (DJF, MAM, JAJ, SON).

Observational data



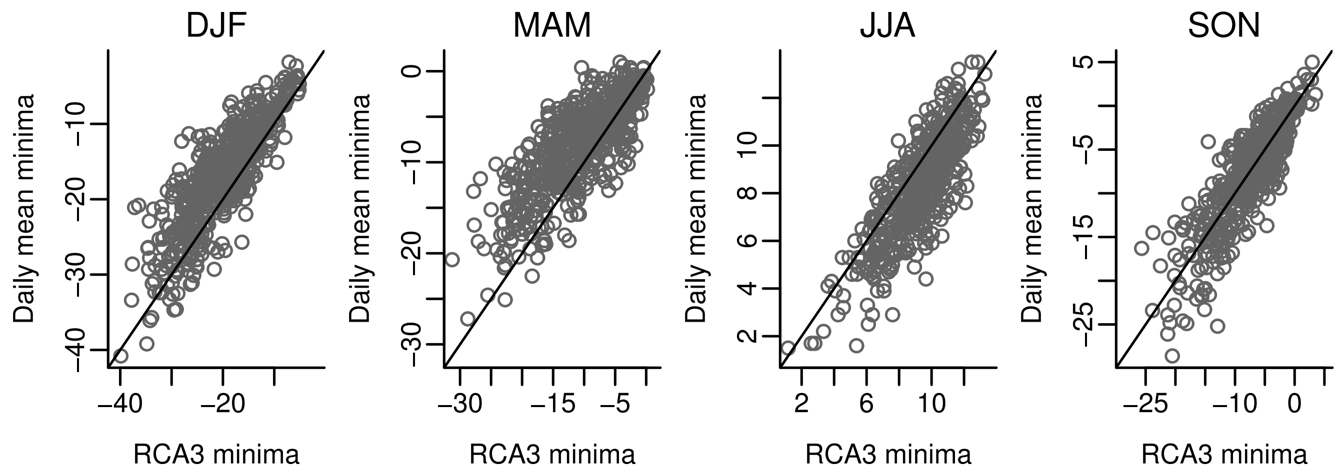
Synoptic observations from 17 sites in an area of south central Sweden

(Also from Swedish Meteorological and Hydrological Institute, SMHI).

Calculate **seasonal minima** (DJF, MAM, JAJ, SON) of

daily mean temperature.

Non spatio-temporal comparisons



Seasonal minima for the station data and RCM agree best in the autumn (SON).

In the winter (DJF) and spring (MAM), the observed minima tend to be slightly higher than is observed for the RCM.

For the summer (JJA), this relationship is reversed – we observe cooler minima than is predicted by the model.

Investigating shifts in the distribution

To compare distributions, we use **Doksum's shift function** [Doksum, 1974].

We find a shift function $\Delta(\cdot)$ such that for random variables

$$X \text{ (RCM) and } Y \text{ (observations)}$$

we have $X + \Delta(X) \sim Y$.

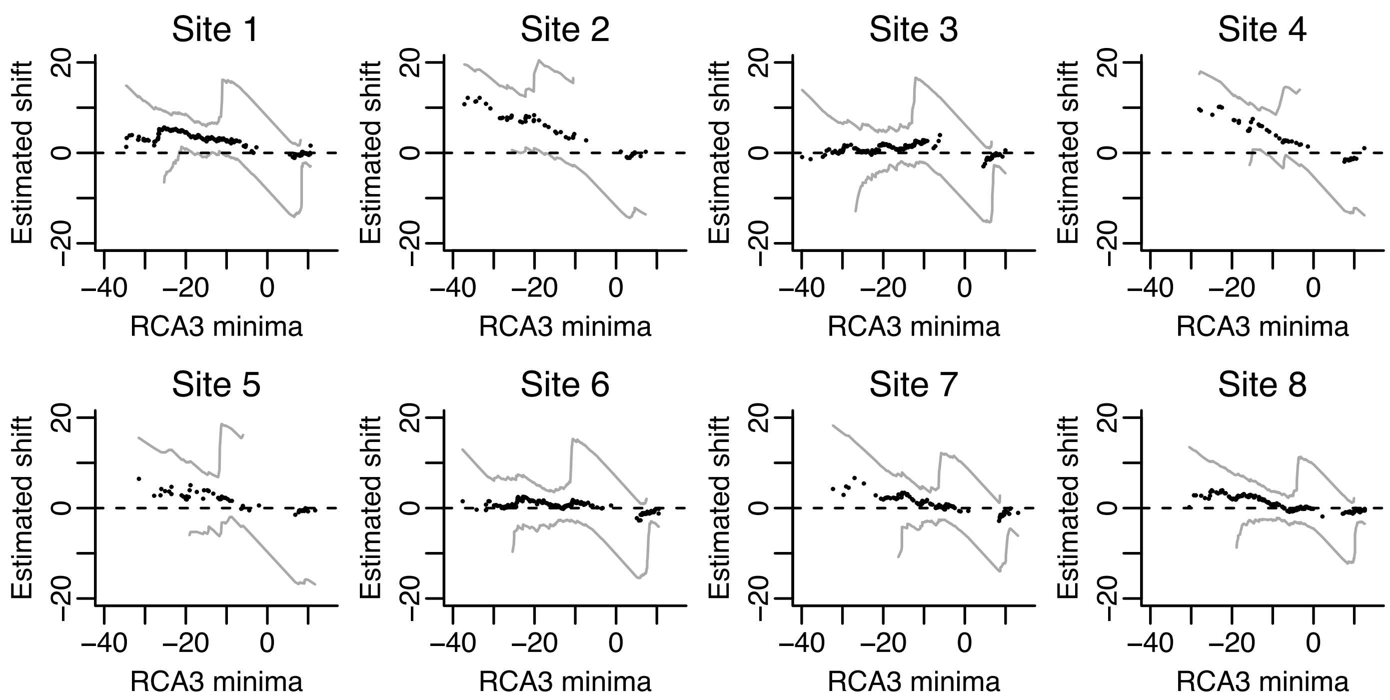
Our **estimate** of $\Delta(x)$ is

$$\widehat{\Delta}(x) = G_m^{-1}(F_n(x)) - x,$$

where F_n (resp. G_m) is the empirical distribution function of X (resp. Y)

Can also construct simultaneous confidence bands for $\Delta(\cdot)$ [Doksum and Sievers, 1976].

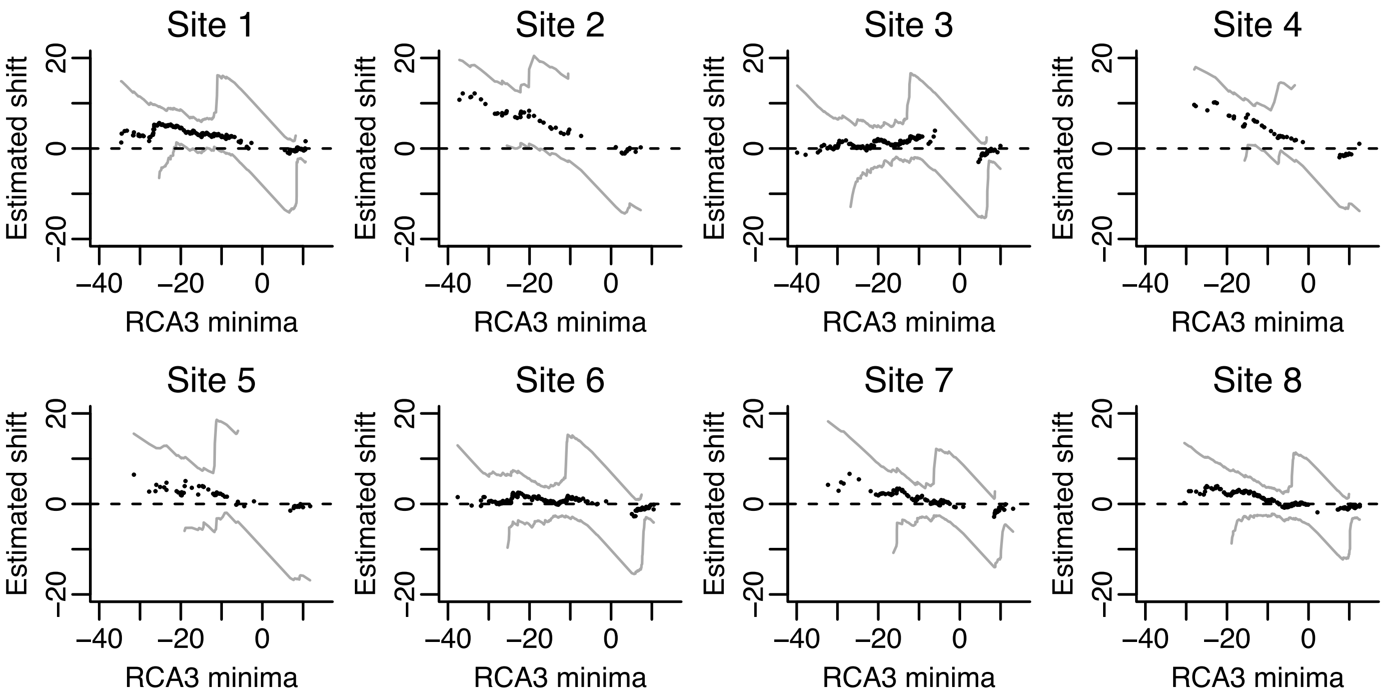
Investigating shifts in the distribution, continued



Anomalous behavior of the RCM around 0°C.

E.g., for grid square containing Station/Site 1 there are no RCA minima between about -2 and 5 °, but there are observed minima in that range of values.

Investigating shifts in the distribution, continued



Minima around 0°C for the RCM are most likely in the autumn (SON).

Less likely for the spring (MAM) and summer, and unlikely in the winter (DJF).

See Nikulin et al. [2011] for an example of the spatial estimation of shifts using the gridded E-OBS data product [Haylock et al., 2008] (with some caveats).

Extreme value theory-based comparisons

Use **(marginal) extreme value theory (EVT)** to analyze and model the seasonal minima for both the **station data** and the **RCM output**.

For example, consider the station data:

At location $\boldsymbol{s} \in D$, let $Z_t(\boldsymbol{s})$ denote the **block minima** in year $y_t(\boldsymbol{s})$ and season $d_t(\boldsymbol{s})$ (taking on values 1: DJF; 2: MAM; 3: JJA; 4: SON), for time index $t = 1, \ldots, N(\boldsymbol{s})$.

Modeling the **negative of the minima** we suppose

$$[-Z_t(\boldsymbol{s})] \sim \mathrm{GEV}(\widetilde{\mu}_t(\boldsymbol{s}), \sigma_t(\boldsymbol{s}), \xi_t(\boldsymbol{s})),$$

conditionally independent over $\boldsymbol{s}$ and t.

The GEV parameters

$\widetilde{\mu}_t(\boldsymbol{s}) \in \mathbb{R}$: **location parameter** indicating values which the distribution of the negative minima are concentrated around.

$\sigma_t(\boldsymbol{s}) > 0$: **scale parameter** defining the spread of the distribution.

$\xi_t(\boldsymbol{s})$: **shape parameter**. The tails of the GEV distributions are heavier for higher values of the shape parameter (A negative shape parameter leads to bounded tails; otherwise the tails of the distribution are unbounded.)

Interpreting the GEV parameters

We think GEV parameters as **describing climate**, with the changes in the parameters indicating seasonal differences and possible trends.

Given the climate, the model technically assumes that weather at different stations is **conditionally independent**.

- A **oversimplification**, since typically events of extremely cold air arise from arctic air moving south during a high pressure situation.
- So given that one station is extremely cold, it is more likely that another is.

We comment further on this at the end!

Modeling the location parameter

Spatio-temporal model for $\{\mu_t(\boldsymbol{s}) = -\widetilde{\mu}_t(\boldsymbol{s})\}$:

$$\mu_t(\boldsymbol{s}) \;=\; \beta_0(\boldsymbol{s}) + \beta_1(\boldsymbol{s})(y_t(\boldsymbol{s}) - 1960) + \sum_{d=2}^{4} \beta_d(\boldsymbol{s}) I(d_t(\boldsymbol{s}) = d).$$

where $I(\cdot)$ is the indicator function.

We assume each $\{\beta_j(\boldsymbol{s})\}$ are independent **Gaussian processes** each with mean λ_j and isotropic covariance, $\mathrm{cov}(\beta_j(\boldsymbol{s}), \beta_j(\boldsymbol{s}+\boldsymbol{h})) = \tau_j \exp(-||\boldsymbol{h}||/\phi_j)$.

Here $\tau_j > 0$ is the (partial) sill parameter, $\phi_j > 0$ is the range parameter, and $||\cdot||$ denotes the Euclidean norm.

Modeling the scale and shape parameters

Assume that the **scale parameters** vary over space, but are constant in time:

$\sigma_t(\boldsymbol{s}) = \sigma(\boldsymbol{s})$ for all t and $\boldsymbol{s}$.

We suppose $\{\log \sigma(\boldsymbol{s})\}$ is a **Gaussian process** with mean λ_σ and isotropic covariance $\mathrm{cov}(\log \sigma(\boldsymbol{s}), \log \sigma(\boldsymbol{s} + \boldsymbol{h})) = \tau_\sigma \exp(-||\boldsymbol{h}||/\phi_\sigma)$.

Our assumption of a **constant shape parameter**, ξ, is an oversimplification, but is reasonable [e.g. Cooley et al., 2007, Sang and Gelfand, 2010].

Bayesian inference

Our **parameters** of interest are

$$\begin{aligned}\boldsymbol{\theta} \;=\; \Big(&\{\beta_j(\boldsymbol{s}) : \boldsymbol{s} \in D, j = 0, \ldots, 4\}, \{\log \sigma(\boldsymbol{s}) : \boldsymbol{s} \in D\}, \xi, \\ &\{\lambda_j : j = 0, \ldots, 4\}, \{\tau_j\}, \{\phi_j\}, \lambda_\sigma, \tau_\sigma, \phi_\sigma\Big)^T.\end{aligned}$$

With the exception of the hyperparameters for the **shape** parameter ξ and the **spatial range** parameters, we assume **vague priors**.

For the **range** parameters we use the gamma prior choice of Craigmile and Guttorp [2011], who modeled daily mean temperature from the same synoptic stations.

The posterior

The **posterior distribution** of $\boldsymbol{\theta}$ given the data is not available in closed form.

We use a Markov chain Monte Carlo (MCMC) algorithm to sample from the posterior distribution.

The algorithm used is adapted from Mannshardt et al. [2013].

Fit one model to the observational data; another model to the RCM output.

The results are **robust** to **minor changes** in this choice of prior distribution.

Approximations required to fit the RCM model

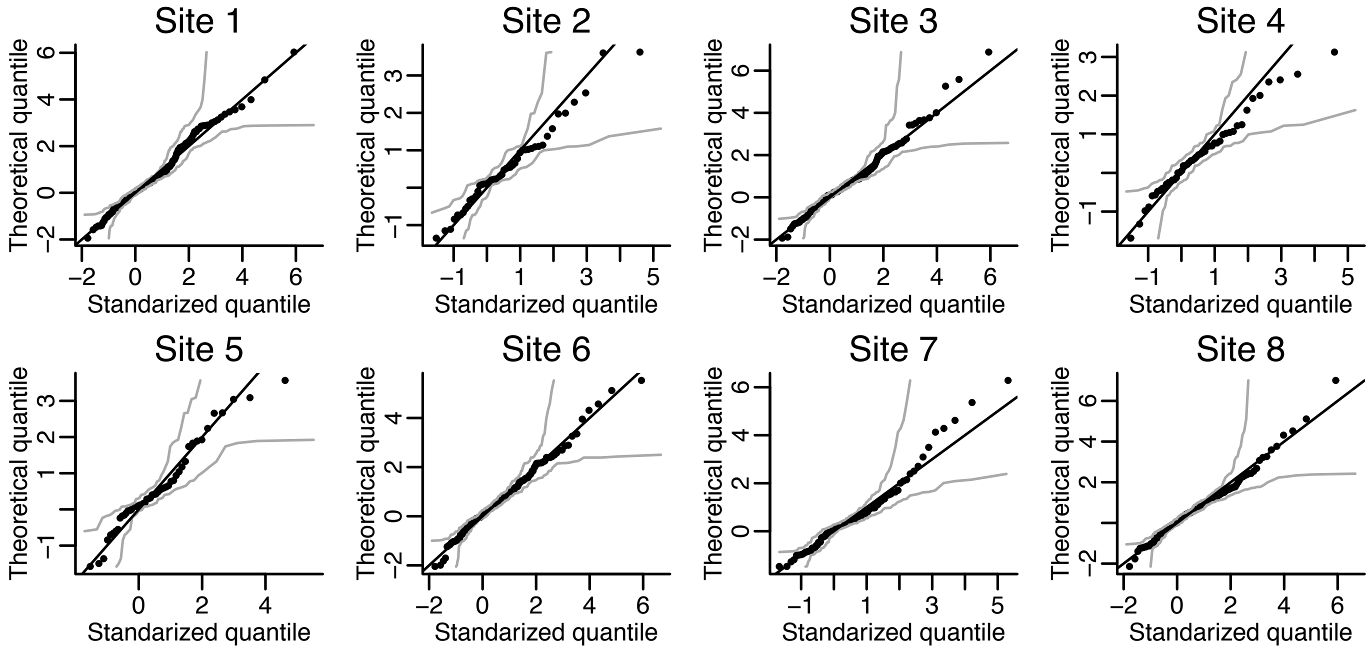
Because we have 1989 spatial locations, we made two approximations:

1. In updating the spatially varying GEV parameters, we calculated the acceptance ratios for Metropolis updates at each spatial location using the 15 **nearest neighbors**, rather than all the spatial locations.

2. In updating the hyperparameters in the spatial models, we broke the spatial field up into 4 **sub-regions** (NW, NE, SW and SE). This sped up matrix inversions.

Experiments demonstrated our results were robust to these approximations.

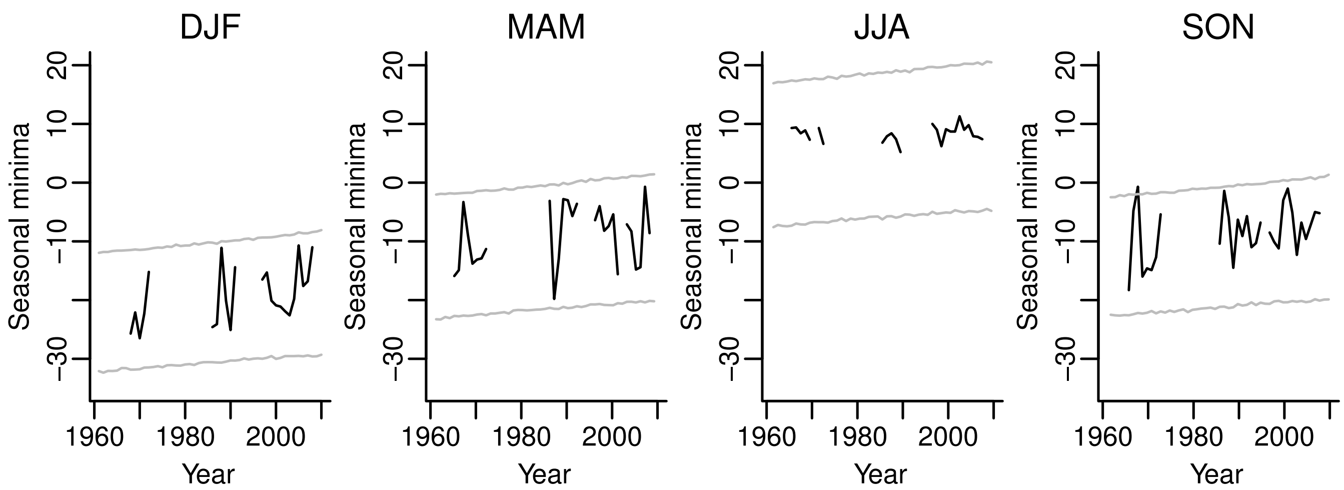
Model verification



Quantile plots [see, e.g., Coles, 2001, Section 6.2.3] were used to assess the distributional assumptions made by the GEV models on a site-by-site basis.

- Excellent goodness of fit at all locations.
- Also indicates model fit improved over a model in which scale parameter was held fixed over locations.

Model verification, cont.



Also assess model using series at Borlänge — located at 15° 30' E and 60° 25' N, at an elevation of 152 meters. Station has been moved twice, and has long stretches of missing data, and was not included in the GEV model fit.

GEV model is **underestimating** the **variation** of seasonal minima in the summer (JJA).

GEV model results: the shape parameters

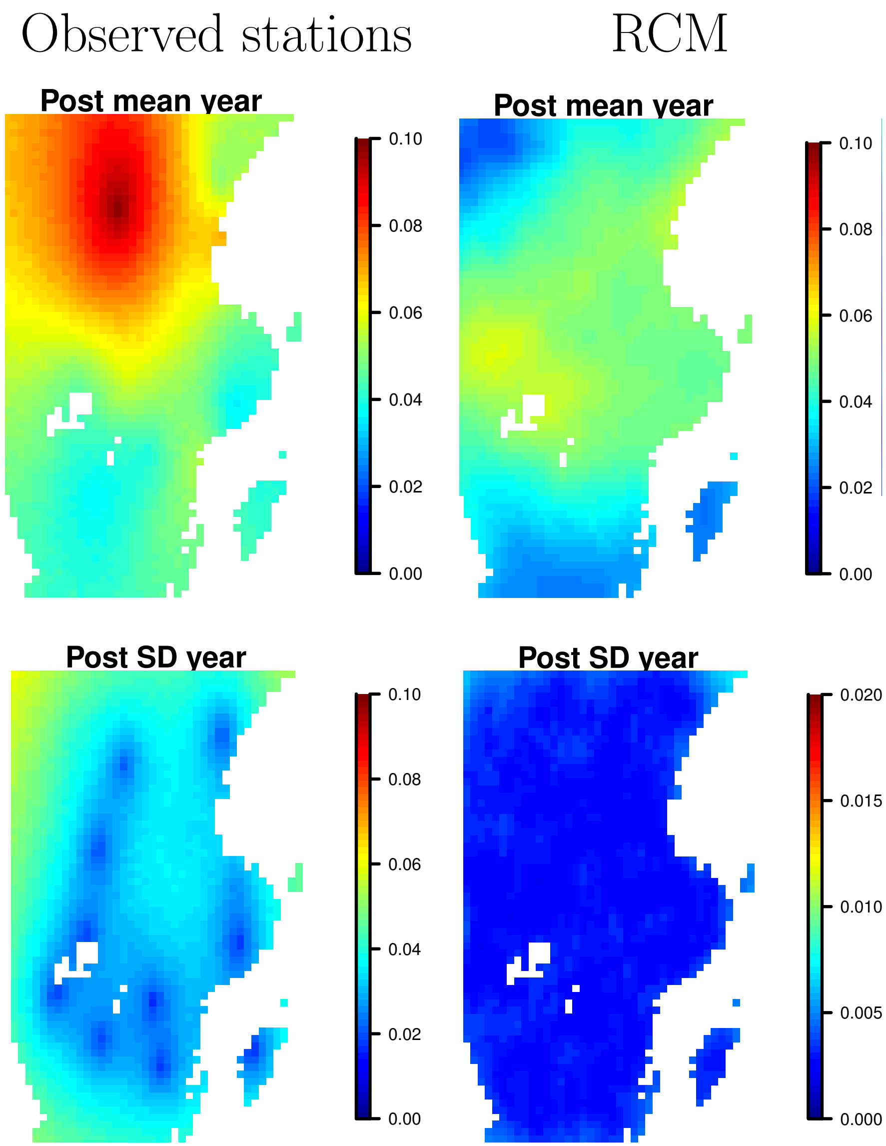
We fit our spatio-temporal GEV model to both the observational data and the RCM output.

	Observed stations		RCM	
Parameter	Post. mean	95% CI	Post. mean	95% CI
ξ	-0.18	(-0.21, -0.15)	-0.14	(-0.15, -0.14)

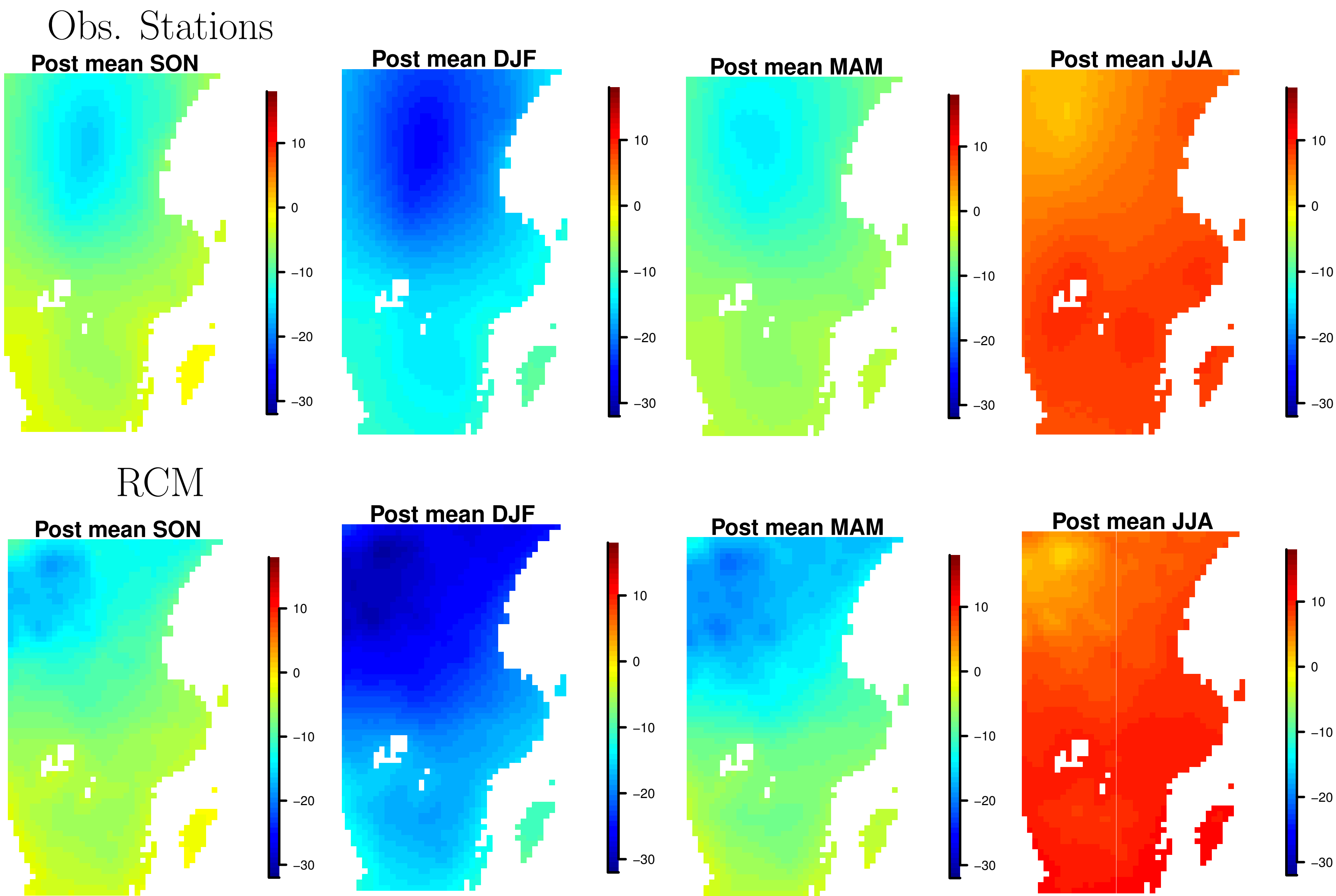
Shape parameters, ζ, are similar in both models.

Both parameters negative: hence distribution of seasonal minima is bounded.

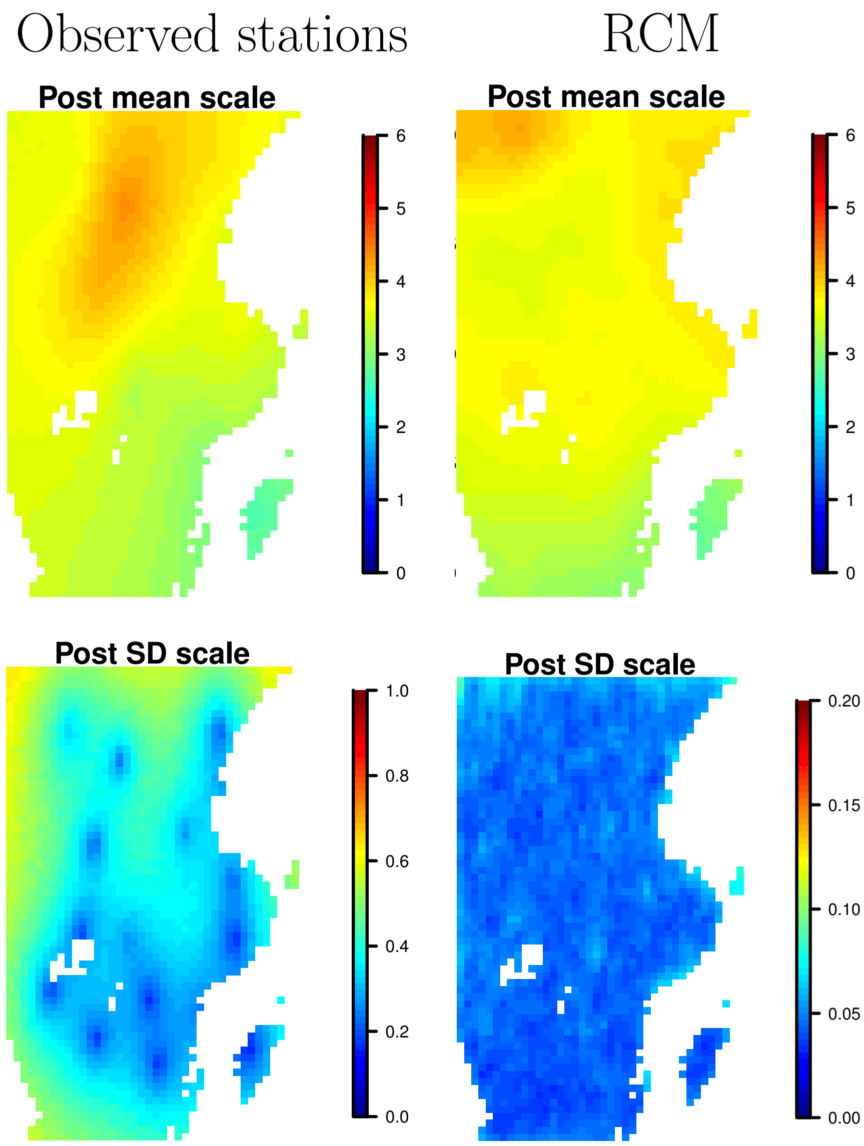
GEV model results: temporal trends



GEV model results: seasonal effects



GEV model results: scale parameters



Summary of GEV modeling results

1. Increasing trend in seasonal minima – warming underestimated in the RCM.

   For observations, change per year ranges from 0.04–0.10 ^oC.

2. Seasonal patterns in observations and RCM output are similar in direction.

   RCM too cold in DJF.

3. Differences in the spatial distribution of the scale parameter.

Thinking about change of support

For RCM output, what does the seasonal minimum calculated for a **grid box** represent?

- Is it the minimum over the grid box region, or the minimum at the centroid?

We could answer these questions by **simulating minima** (based on the observation model) at a **finer** spatial scale than the regional climate model output, and then changing support by calculating minima at the grid box level.

But requires the use of **multivariate** models for extremes...

Other extreme value analyses

We used a block minima GEV approach, assuming **conditional independence** (there are obvious extensions to the GEV models we use).

Would be interesting to compare to a **threshold** approach.

More important: **joint modeling of extremes**

- See all the other wonderful talks at this workshop!

Other extreme value analyses, cont.

Number of useful approaches worth investigating:

Conditional modeling [e.g., Ledford and Tawn, 1996, Heffernan and Tawn, 2004]

Max-stable approach [e.g. Davison et al., 2012]

Copula-based approaches [e.g., Ghosh and Mallick, 2011]

Multivariate inference is computationally demanding [e.g. Ribatet et al., 2012].

Bias correction

We are thinking more about bias correction methods for RCM output [e.g., Bordoy and Burlando, 2013].

References

R. Bordoy and P. Burlando. Bias correction of regional climate model simulations in a region of complex orography. *J. Appl. Meteor. Clim.*, 52:82–101, 2013.

S. Coles. *An Introduction to Statistical Modeling of Extreme Values*. Springer-Verlag, 2001.

D. Cooley, D. Nychka, and P. Naveau. Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association*, 102: 824–840, 2007.

P. F. Craigmile and P. Guttorp. Space-time modelling of trends in temperature series. *Journal of Time Series Analysis*, 32:378–395, Apr. 2011.

A. C. Davison, S. A. Padoan, and M. Ribatet. Statistical modelling of spatial extremes. *Statistical Science*, 27:161–186, 2012.

K. A. Doksum. Empirical probability plots and statistical inference for nonlinear models in the two sample case. *Ann. Statist.*, 2:267–277, 1974.

K. A. Doksum and G. L. Sievers. Plotting with confidence: Graphical comparisons of two populations. *Biometrika*, 63:421–434, 1976.

S. Ghosh and B. K. Mallick. A hierarchical Bayesian spatio-temporal model for extreme precipitation events. *Enivronmetrics*, 22:192–204, 2011.

M. R. Haylock, N. Hofstra, A. M. G. Klein Tank, E. J. Klok, P. D. Jones, and co authors. European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. *Journal of Geophysical Researh*, 113, 2008. doi: 10.1029/2008JD010201.

J. E. Heffernan and J. A. Tawn. A conditional approach for multivariate extreme values. *Journal of the Royal Statistical Society, Series B*, 66:497–546, 2004.

A. W. Ledford and J. A. Tawn. Statistics for near independence in multivariate extreme values. *Biometrika*, 83:169–187, 1996.

E. Mannshardt, P. F. Craigmile, and M. P. Tingley. Statistical modeling of extreme value behavior in North American tree-ring density series. *Climatic Change*, 117:843–858, 2013.

G. Nikulin, E. Kjellström, U. Hansson, G. Strandberg, and A. Ullerstig. Evaluation and future projections of temperature, precipitation and wind extremes over Europe in an ensemble of regional climate simulations. *Tellus A*, 63:41–55, 2011.

M. Ribatet, D. Cooley, and A. C. Davison. Bayesian inference from composite likelihoods, with an application to spatial extremes. *Statistica Sinica*, 22:813–845, 2012.

P. Samuelsson, C. G. Jones, U. Willén, A. Ullerstig, S. Gollvik, U. Hansson, C. Jansson, E. Kjellström, G. Nikulin, and K. Wyser. The Rossby Centre regional climate model RCA3: model description and performance. *Tellus A*, 63:4–23, 2011.

H. Sang and A. E. Gelfand. Continuous Spatial Process Models for Spatial Extreme Values. *Journal of Agricultural, Biological and Environmental Statistics*, 15: 49–65, 2010.

S. Uppala, P. Kölberg, A. Simmons, U. Andrae, V. da Costa Bechtold, M. Fiorino, J. Gibson, J. Haseler, A. Hernandez, G. Kelly, X. Li, K. Onogi, S. Saarinen, N. Sokka, R. Allan, E. Andersson, K. Arpe, M. Balmaseda, A. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, S. Caires, F. Chevallier, A. Dethof, M. Dragosavac, M. Fisher, M. Fuentes, S. Hagemann, E. Hólm, B. Hoskins, L. Isaksen, P. Janssen, R. Jenne, A. McNally, J.-F. Mahfouf, J.-J. Morcrette, N. Rayner, R. Saunders, P. Simon, A. Sterl, K. Trenberth, A. Untch, D. Vasiljevic, P. Viterbo, and J. Woollen. The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society*, 131:2961–3012, 2005.

S. Uppala, D. Dee, S. Kobayashi, P. Berrisford, and A. Simmons. Towards a climate data assimilation system: status update of ERA-Interim. *ECMWF Newsletter*, pages 12–18, 2008.