Improved 21st century projections of sub-daily extreme precipitation by spatio-temporal recalibration

Specific aims of the project, research questions/hypotheses, feasibility, originality and impact

This project will provide more robust projections of future changes in local extremes by developing and applying a new statistical framework that combines for the first time high-resolution climate model outputs and historical observations in a joint model, providing unprecedented information on extreme sub-daily precipitation. The project will deliver this statistical method and bias-corrected projections of extreme precipitation across the UK on a 5 x 5 km grid for 2021-2040 and 2061-2080, and these will be made easily accessible for decision makers, stakeholders and others. As an important output, the project will improve predictions of intense hourly precipitation. These predictions are crucial for understanding future changes in flash floods that are one of the costliest and most critical of climate change impacts, and are known to be the cause of fatalities, water resource and transportation disruption, and cause damage to infrastructure and ecosystems.

The project will result in: (i) a computationally feasible statistical model that scales well with the spatial dimension. It is based on a hierarchical structure equipped with a copula at the data level, able to take account of extremal spatial dependence and time-varying covariates, and designed to handle block maxima. Its inference scheme is an extension of the recently developed Max-and-Smooth method (Hrafnkelsson et al., 2021). The project will also provide: (ii) freelyavailable R language computer code for the method; (iii) bias-corrected projections of annual and seasonal maximum 1-h, 3-h, 6-h, 12-h and 24-h precipitation on a fine grid across the UK over the periods 2021-2040 and 2061-2080; (iv) estimates of quantiles and their uncertainty for the variables in (iii); (v) user-friendly interface to access these quantities; (vi) statistical generator for the variables in (iii). Outputs (i) and (ii) will benefit the wider research community, and the products in (iii), (iv), (v) and (vi) will be useful for policymakers and stakeholders as an input for flood risk quantification in a changing climate.

The project will address two main objectives. The first objective is development of: (i) a joint statistical model for two sources of extreme data on a high dimensional grid that can simultaneously handle extremal spatial dependence and time-varying covariates; (ii) a suitable inference scheme for the model parameters that efficiently handles high spatial dimensions; and (iii) an R package for this model and its inference scheme. The second objective involves: (i) applying the joint statistical model and its inference scheme to extreme precipitation observations on a sub-daily scale and an output from a convection-permitting model from the United Kingdom Meteorological Office (UKMO), both available on a fine (5 km) high dimensional grid; (ii) using the output of the statistical model to produce bias-corrected projections for annual and seasonal maximum precipitation (1-h, 3-h, 6-h, 12-h, 24-h) for 2021-2040 and 2061-2080 and quantiles of sub-daily extremes for each grid point and each year in the period 2021-2080, and (iii) making them easily accessible to end users, such as decision makers, regulators, consultants and other stakeholders, via a user-friendly web-based interface developed within the project.

This project is feasible for several reasons: (i) the UKMO launched a net set of climate projections in 2019 (UKCP Local), which were updated in July 2021, at small spatial and temporal scales (5 x 5 km and 1-h) (Kendon et al., 2021); (ii) a quality-controlled hourly precipitation dataset for the UK from 1992-2014 on an hourly scale and on a 1 x 1 km spatial grid is available and thus compatible to the UKMO output. Analysis based on the projections revealed that present-day return levels of 1-h precip. extremes are overestimated by 20-25% (Kendon et al., 2020a), showing that recalibration is feasible. Two of the team members behind this project have full access to the data in (i) and (ii) and they are ready for use; (iii) a statistical method for handling multi-parameter distributions such as the generalized extreme value (GEV) distribution on large spatial scales has been recently developed by three members of the project team, and it can be extended for simultaneously handling the extremal spatial dependence and time varying covariates, which are features of the extreme precipitation output and the historical observations of extreme precipitation; (iv) decision makers and stakeholders (Environment Agency; SEPA; UK Centre for Ecology and Hydrology (UKCEH); UK Water companies; Network Rail; Insurance companies) will be invited to a workshop to learn about the project and to discuss what outputs should be made publicly available and what form these should take. One of the team members behind the project has extensive experience in organizing these workshops and key contacts in these organisations already; (v) a team of six experts has been assembled, covering the fields of; statistical modeling and computation; extreme value theory; and climate science and modeling.

The originality of the project lies in coupling the sub-daily precipitation extremes on a fine grid from the convection-permitting climate model with observed sub-daily precipitation extremes on the same grid and with overlap in time, and the development of a statistical method that can handle the extremal spatial features and the high dimensionality of these two datasets. This is the first time bias-corrected projections for extreme precipitation on fine space (5 km) and time (sub-daily) scales will be made available to end users and others. This type of information is crucially needed for decision making around future flood risks.

The methods that will be developed for the UK data will be applicable to data with the same structure from other areas across the World, including Iceland when data become available. Currently, there is an EU effort on climate projections with high resolution in space and time (CORDEX-FPS and EUCP projects). The statistical community will welcome this methodological development as the computational cost of the current methods for extremal spatial dependence do not scale as well as the proposed method with increasing spatial dimension.

The scientific value and the societal impact of the project lies in the following. First, the recently launched projections of extreme sub-daily precipitation across the UK are valuable due to their high resolution in space (5 km) and time (1 hour) and the fact that the convection-permitting model gives a much more realistic representation of atmospheric convection, and is able to capture the characteristics of hourly rainfall, including extremes, unlike coarser resolution climate models. Due to the small grid-scale it also better represents topographic effects.

A bias-corrected version of these projections will be invaluable as inputs into various impact models for decision makers and other end users to help with adaptation planning and climate resilience management, and it will be made easy for them to access. Furthermore, the projections stretch over 2021-2040 and 2061-2080, thus, bias-corrected projections of extreme sub-daily precipitation in a changing climate will be provided. Second, posterior predictions of sub-daily extreme precipitation from 2020 to 2080 on the same fine spatial scale (5 km) along with posterior uncertainty will be provided. These predictions provide an alternative way to present the bias-corrected projections and are also useful for decision making. Third, a statistical model and a corresponding inference scheme will be developed. The data consist of both the projections from the high-resolution climate model and historical observational data. Our statistical framework will model these two data sources jointly and account for both the spatial dependence found in the data and a time varying covariate. The inference scheme will be designed to scale well with the spatial dimension of the data, and it can easily be adjusted to handle sparse observational data, but it requires climate model output on a grid, preferably a high resolution grid. Finally, the proposed statistical method is general since it can be applied to other climate variables, e.g., maximum and minimum temperatures, and maximum wind speed.

Present state of knowledge in the field

In 2018 the UKMO launched climate projections referred to as the UKCP18 climate projections. The 2019 update of the UKCP climate projections were the first ones worldwide with resolution on a par with operational weather forecast models (Kendon et al., 2020b). These climate projections can be used for national climate scenarios and allow for examining the risk of extreme weather events at a local scale over the next decades. Previous studies, e.g., Kendon et al. (2014), show that convection-permitting climate models, such as those used within UKCP, can represent 1-h precipitation and its extremes, much more realistically than outputs from conventional climate models at coarser spatial scales. Currently, the UKCP climate projections contain data across the UK over three 20-year periods (1981-2000, 2021-2040 and 2061-2080) under a high-emission scenario RCP8.5. An ensemble of 12 members, driven by the Strand 3 12 km RCM ensemble, is available. Elizabeth Kendon, team member, has full access to these data.

The historical observations of the gridded sub-daily extreme precipitation are based on an hourly precipitation dataset that is derived from precipitation gauges in the UK covering the years from 1992-2014, see Blenkinsop et al. (2017) and Lewis et al. (2018). The part of the dataset covering 1992-2011, was gathered by Blenkinsop et al. (2017). It comes from three sources: (i) the UK Met Office Integrated Data Archive System (MIDAS); (ii) the Scottish Environmental Protection Agency (SEPA); and (iii) the UK Environment Agency (EA). In Blenkinsop et al. (2017) a number of site-specific quality control procedures were applied to the data to detect accumulated totals, malfunctioning gauges and unfeasible extreme precipitation totals. The dataset was extended to 2014 and additional quality control checks against neighboring gauges were performed (Lewis et al., 2018). Recently, the additional criteria of having at least 85% of the gauge record complete (i.e., non-missing and data not flagged by the QC process) was added for each of the years in the period 1992-2014 to ensure having complete and representative data (Darwish et al., 2020). A total of 197 gauges distributed across the UK fulfilled this criterion. One of the team members, Hayley Fowler, has full access to these data.

Bayesian hierarchical models (BHM) have been used to model spatial extremes over the last two decades. BHMs consist of three levels (data level, latent level, hyperparameter level) and they model the data and the parameters at each level with probability distributions. This framework has been very useful in many applications. In this project the block maxima approach will be applied to the precipitation extremes, partially, since that facilitates scalable inference scheme, see below. These models have a clear structure, however, they often lead to a reduced amount of information in comparison with peaks over threshold models (de Haan and Ferreira, 2006; Cooley and Sain, 2010). The block maxima approach has been applied to extreme precipitation; see, e.g., Geirsson et al. (2015) and Dyrrdal et al. (2015). In the last three papers above, the BHMs are latent Gaussian models (LGMs) since the prior density for the latent parameters is Gaussian, furthermore, the data are modeled independently conditional on the model parameters. In the case of precipitation data, this conditional independence assumption is poor because some spatial dependence exists at the data level due to the proximity between stations. Even when the dependence is not of interest and the main interest is in the marginal distributions at each site, incorrectly assuming conditional independence yields sub-optimal results since the uncertainty in the model parameters is underestimated. Thus, to improve uncertainty quantification, the spatial dependence at the data level needs to be accounted for. Several models have been proposed for modeling extremal spatial dependence simultaneously, e.g., the Gaussian-copula, the t-copula and max-stable processes (e.g., Reich and Shaby, 2012).

Hrafnkelsson et al. (2021) proposed a method referred to as Max-and-Smooth for inference in latent Gaussian models with a multivariate link function. Max-and-Smooth is an approximate Bayesian inference scheme for LGMs set up such that one or more of the likelihood parameters are modeled by latent additive Gaussian processes. This inference scheme is based on two-steps. In the first step (Max), the likelihood function is approximated by a Gaussian density with mean equal to the maximum likelihood estimate and covariance equal to the inverse observed information. An alternative approach entails using the mean and covariance of the normalized likelihood function. Johannesson et al. (2021) showed that the Gaussian approximation to the GEV likelihood is reasonably good for moderate sample sizes, i.e., 20 observations or more per site. Here the sample size is 23 and 60 in the case of the observations and the projections, respectively. The latent parameters and hyperparameters are inferred within the second step (Smooth) by applying the approximated likelihood function along with the prior densities of these parameters. Max-and-Smooth ensures that the uncertainty from the first step is correctly propagated to the second step and it is fully Bayesian.

The approximated posterior density of the latent parameters (conditional on the hyperparameters) is Gaussian due to the Gaussian form of both the approximated likelihood function and the prior density. As a result posterior computation in high dimensions becomes efficient. The hyperparameters can be sampled independently of the latent parameters since their approximate marginal posterior distribution is tractable. This is known to give more efficient posterior samples compared to Gibbs sampling where the hyperparameters are sampled conditionally on the latent parameters (Filippone et al., 2013). The computational cost of Max-and-Smooth is close to being insensitive to the number of independent data replicates. Furthermore, when the Gaussian prior density is specified with a sparse precision matrix, Max-and-Smooth scales well with increased dimension of the latent parameter vector. Max-and-Smooth is applicable to the wide class of extended LGMs, and it is relatively easy to implement. In particular, in the case of extended LGMs for spatio-temporal datasets with a large number of data replicates per cell, modeled with high-dimensional latent vectors and sparse precision matrices, the speedup is substantial in comparison to an MCMC scheme that infers the posterior density from the exact likelihood function.

Previous studies show that extreme precipitation exhibits spatial dependence (e.g., Bopp et al., 2021). The purpose of the statistical model and its inference scheme, designed for the UK extreme precipitation data, includes capturing the uncertainty of estimated model parameters and generating extreme values from it. To achieve this and to have a computationally feasible inference scheme, we opt for extending the Max-and-Smooth approach in Hrafnkelsson et al. (2021) by describing the spatial dependence structure at the data level of the LGM with a t-copula that specifies spatial dependence with a sparse precision matrix as that would allow for the extended Max-and-Smooth approach to scale well with the dimension of the latent parameters.

In Lehmann et al. (2016), extreme precipitation is analyzed using a climatological covariate at the latent level for each of the three parameters of the generalized extreme-value distribution at each site. This climatological covariate consists of estimates of the GEV parameters at each observational site that are based on the output of a Regional Climate Model (RCM) on a grid. To model precipitation extremes for various duration lengths, Lehmann et al. (2016) used an empirical model from Koutsoyiannis et al. (1998), in which the GEV parameters are a function of the duration, d, and a few other parameters. The models in these two papers will be used as references when modeling the UK extreme precipitation data.

Management and co-operation

The project will be hosted at the University of Iceland. It will be led by a team of six experts, the PI, Birgir Hrafnkelsson, and five co-proposers, see below. Each of the experts will advise the PhD student and the postdoctoral researcher with the various aspects of the project. The PI will be the PhD student's main advisor. The PhD student and the postdoctoral researcher will have access to all the needed computing facilities to successfully finish the project. They will also have full access to the UKMO climate projections and the observed hourly precipitation dataset for the UK through the team members Elizabeth Kendon at UKMO and Hayley Fowler at Newcastle University, respectively.

An external advisory board will be set up for this project. It will consist of the PI and the five co-proposers. To help develop the work and interact with the co-proposers, the PhD student and the postdoctoral researcher will visit the co-proposers. The PhD student and the postdoctoral researcher will visit the UKMO and University of Exeter, both in Exeter, once every year for 1-2 weeks. Furthermore, the postdoctoral researcher will be hosted at UKMO during the last 6-12 month out of 36 months. The PhD student and the postdoctoral researcher will also visit Newcastle University and King Abdullah University of Science and Technology (KAUST).

The team members have worked together in pairs or triplets on various projects and papers in the past. Together the members of the team cover the expertise required by the project, namely, extreme value theory, spatial statistics, Bayesian hierarchical modeling and computation, climate models, statistics for climate sciences. The team members are:

Principal Investigator and main PhD advisor: Birgir Hrafnkelsson is Prof. of Statistics at University of Iceland. He is an expert in computation for LGMs, spatial statistics, spatio-temporal models, and applications of LGMs to the environmental sciences and geophysics.

Co-proposer: Hayley Fowler is Prof. of Climate Change Impacts at Newcastle University, UK. She is a Fellow of the AGU and an expert in extreme rainfall, particularly at sub-daily scales. She leads the GEWEX sub-daily precipitation cross-cut internationally.

Co-proposer: Raphaël Huser is an Assistant Prof. of Statistics at the King Abdullah University of Science and Technology (KAUST), Saudi Arabia. He is an expert in statistics of extremes, spatio-temporal statistics, copula modeling, and environmental statistics. He leads the Extreme Statistics Research Group at KAUST.

Co-proposer: Elizabeth Kendon is Manager of Understanding Regional Climate Change at the UKMO Hadley Centre and Prof. at Bristol University. She is an expert in the field of convection-permitting climate modeling (Kendon et al., 2014; Kendon et al., 2019a), and led the design and delivery of the UKCP convection-permitting climate scenarios (Kendon et al., 2019b).

Co-proposer: Stefan Siegert is a Lecturer in Statistical Science at the University of Exeter. He is an expert in spatio-temporal modeling of environmental data, statistical post-processing of weather and climate forecasts, and statistical methodology for forecast quality assessment. He has developed publicly available R packages interactive web tools, and web sites.

Co-proposer: Prof. David Stephenson is the founder and Director of Exeter Climate Systems and has 25 years experience in climate modeling research. He is world-recognized for his expertise in the development and novel application of statistical modeling to understand climate processes and predictions. He has led statistical work in several recent projects e.g. EU SPECS, Copernicus QA4Seas, NERC EuroCLIM.

References

Bopp, G., Shaby, B., and Huser, R. (2021), A hierarchical max-infinitely divisible spatial model for extreme precipitation. *Journal of the American Statistical Association (Applications and Case Studies)* **116**, 93-106.

Blenkinsop, S., Lewis, E., Chan, S.C. and Fowler, H.J. (2017) Quality-control of an hourly rainfall dataset and climatology of extremes for the UK. *International Journal of Climatology*, 37(2), 722-740.

Cooley D. and Sain S.R. (2010). Spatial Hierarchical Modeling of Precipitation Extremes from a Regional Climate Model. *Journal of Agricultural, Biological, and Environmental Statistics*, 15, 381-402.

Darwish, M., Tye, M., Prein, A. Fowler, H.J., Blenkinsop, S. (2020) Statistical modelling of UK extreme hourly precipitation and future climate change responses. In preparation.

de Haan, L. and Ferreira, A. (2006) *Extreme Value Theory.* Springer Series in Operations Research and Financial Engineering. Springer, New York.

Dyrrdal, A. V., Lenkoski, A., Thorarinsdottir, T. L. and Stordal, F. (2015). Bayesian hierarchical modeling of extreme hourly precipitation in Norway. *Environmetrics*, 26(2), 89-106.

Filippone, M., Zhong, M., and Girolami, M. (2013). A comparative evaluation of stochastic-based inference methods for Gaussian process models. *Machine Learning*, 93(1): 93-114.

Geirsson, O.P., Hrafnkelsson, B. and Simpson, D. (2015) Computationally efficient spatial modeling of annual maximum 24-h precipitation on a fine grid. *Environmetrics*, 26(5), 339-353.

Hrafnkelsson, B., Siegert, S., Huser, R., Bakka, H. and Johannesson, A.V. (2021) Max-and-Smooth: A two-step approach for approximate Bayesian inference in latent Gaussian models. *Bayesian Analysis*, 16(2), 611-638.

Johannesson, A.V., Siegert, S., Huser, R., Bakka, H. and Hrafnkelsson, B. (2021) Approximate Bayesian inference for analysis of spatio-temporal flood frequency data. *arXiv preprint* arXiv:1907.04763v3.

Kendon, E.J., Roberts, N.M., Hayley J. Fowler, H.J., Malcolm J. Roberts, M.J., Steven C. Chan, S.C., and Senior, C.A. (2014) Heavier summer downpours with climate change revealed by weather forecast resolution model. *Nature Climate Change*, 4, 570-576.

Kendon, E.J., R.A Stratton, S. Tucker, J.H. Marsham, S. Berthou, D.P. Rowell and C. A. Senior (2019a) Enhanced future changes in wet and dry extremes over Africa at convection-permitting scale. *Nature Communications*.

Kendon, E.J., et al. (2019b) UKCP Convection-permitting model projections: Science report, UK Met Office Hadley Centre, Exeter. Open access:
https://www.metoffice.gov.uk/pub/data/weather/uk/ukcp18/science-reports/UKCP-Convection-permitting-model-projections-report.pdf.

Kendon, E.J., et al. (2020a) Technical note: Issue with UKCP Local (2.2km) simulation data, UK Met Office Hadley Centre, Exeter. Open access:
https://www.metoffice.gov.uk/binaries/content/assets/metofficegovuk/pdf/research /ukcp/ukcp_graupel_technical_note.pdf.

Kendon E.J., Roberts, N.M., Fosser, G., Martin, G.M., Lock, A.P., Murphy, J.M., Senior, C.A. and Tucker, S.O. (2020b) Greater future UK winter precipitation increase in new convection-permitting scenarios. *Journal of Climate*. DOI: 10.1175/JCLI-D-20-0089.1

Kendon, E.J., et al. (2021) Update to UKCP Local (2.2km) projections. UK Met Office Hadley Centre, Exeter. Open access:
https://www.metoffice.gov.uk/pub/data/weather/uk/ukcp18/science-reports /ukcp18_local_update_report_2021.pdf

Koutsoyiannis D., Kozonis D. and Manetas A. (1998) A mathematical framework for studying rainfall intensity-duration-frequency relationships. *Journal of Hydrology*, 206(1), 118-135.

Lehmann, E.A., Phatak, A., Stephenson, A. and Lau, R. (2016) Spatial Modelling Framework for the Characterisation of Rainfall Extremes at Different Durations and under Climate Change. *Environmetrics*, 27(4), 239-251.

Lewis, E., Quinn, N., Blenkinsop, S., Fowler, H.J., Freer, J., Tanguy, M., Hitt, O., Coxon, G., Bates, P. and Woods, R. (2018) A rule based quality control method for hourly rainfall data and a 1 km resolution gridded hourly rainfall dataset for Great Britain: CEH-GEAR1hr. *Journal of Hydrology*, 564, 930-943.

Reich, B.J. and Shaby, B.A. (2012). A hierarchical max-stable spatial model for extreme precipitation. *The Annals of Applied Statistics*, 6(4), 1430-1451.