STAGE: STAtistical modelling of Groundwater Extremes

Ben Marchant[a], Emma Eastoe[b], Jenny Wadsworth[b] and John Bloomfield[c]

July 29, 2016

a: British Geological Survey, Keyworth, NG12 5GG, UK
b: Department of Mathematics and Statistics, Fylde College, Lancaster University, Lancaster, LA1 4YF, UK
c: British Geological Survey, Maclean Building, Crowmarsh Gifford, Wallingford, Oxfordshire, OX10 8BB, UK

1   Summary

The primary aim of this feasibility study was to establish a sustainable collaborative relationship between the British Geological Survey (BGS) and the Modelling and Inference Group at Lancaster University (LU) with particular focus on the study of extreme events in the earth sciences. This was achieved through:

- A case study exploring extreme events in long-term (> 100 years) groundwater hydrographes from UK aquifers.

- Networking events designed to introduce BGS earth scientists to LU statisticians and to identify research areas and funding opportunities for future collaborations.

Through the course of the case study, BGS researchers have been introduced to state-of-the-art extreme value statistics methodologies and have been provided with the knowledge and software required to replicate these analyses on other datasets. These methodologies are considerably more flexible and generally applicable than those that currently appear in the hydrological literature. BGS continue to apply these methodologies to groundwater data within the NERC funded 'Analysis of historic drought and water scarcity in the UK' and other internally funded projects concerned with the drivers of groundwater levels. Once the methodologies have been tested on a wider range of datasets, BGS and LU envisage a joint publication assessing the suitability of extreme value statistics to predict the chacteristics of extreme groundwater events. BGS have included extreme

value techniques in a bid for funding from the NERC Knowledge Exchange Programme and in the IMPOWER (Impact for urban groundwater management program) proposal to a consortium of representatives of the water industry. BGS and LU will seek to collaborate in any future programs concerned with the occurrence and risk of floods and droughts in the UK.

The networking activities identified several further topics that are suitable for future collaborations including space-weather, volcanic eruptions, the retreat of glaciers, seismology and soil contamination. These discussions have led to three MSc projects that are currently being conducted by LU students in collaboration with BGS scientists and further project proposals are being developed for the 2016-17 academic year. The project team see these MSc projects as an ideal opportunity for the LU statisticians to make an initial assessment of the available data and the potential for further work and applications for external funding. In each topic area, potential sources for future funding have been identified. BGS and LU are producing a PhD proposal on the spatial prediction of soil properties for submission to the Soils Training and Research Studentships Doctoral Training Centre in October 2016.

2   Networking Activities

The following networking activities were conducted during the feasibility project:

2.1   Initial Meeting

A introductory project meeting at Lancaster University attended by Dr Emma Eastoe (Lancaster University Statistician), Dr Jenny Wadsworth (Lancaster University Statistician), Dr Ben Marchant (BGS Environmental Statistician) and Dr John Bloomfield (BGS Hydrogeologist). This was the first meeting between the primary participants in the project. The main outcome of the meeting was an agreed plan of action for the case study on extreme groundwater levels. Through a series of powerpoint presentations the LU statisticians were introduced to the problems of groundwater floods and droughts and the climatic drivers of these extreme events. The available groundwater level datasets were discussed and it was agreed that the case study would focus on time series data from two sites - Chilgrove House and Dalton Holme - where more than 100 years of monthly groundwater level observations were available. Monthly rainfall and potential evapotranspiration data were also available for these sites. The LU statisticians presented some of the exploratory analysis methodologies that they could apply to such time series. It was agreed that the LU statisticians would conduct initial analyses of the data which would focus on (i) the relationships between the climatic covariates and the occurrence and severity of extreme groundwater events (ii) the characteristics of the extremal dependence of these time series. BGS would provide groundwater expertise, advice and feedback during the course of this work. The results of the initial analyses were discussed in a second project meeting at Lancaster University on February 6 2016. At this meeting Emma Eastoe, Jenny Wadsworth, and Ben Marchant agreed on the most useful form of groundwater extreme models. This work was completed by LU and the results are presented in Section 4.

2.2   MSc Projects

Ben Marchant consulted with colleagues at BGS who proposed three projects for the Environmental Pathway of the LU MSc in Statistics. These projects were seen as an ideal opportunity to introduce the LU statisticians to wider examples of the data held by BGS, to allow the students to work on problems that were of direct interest to BGS researchers, to demonstrate the value of statistical

modelling to the BGS researchers and for the LU statisticians to make exploratory analyses of the BGS data. The three proposals (see Appendix) were concerned with (i) the occurrence of extreme space weather events; (ii) the relationships between the texture (particle size distribution) of soil and its potential to sequester carbon and (iii) the analysis of high frequency groundwater data and its relationship with local weather data and river levels. Emma Eastoe and Jenny Wadsworth supervised three students, two of whom worked on the space weather project and the third worked on the soil project.

2.3 Interviews with BGS researchers at Keyworth

The project proposal included a workshop to be held at the BGS head office in Keyworth where BGS researchers would be introduced to the potential benefits of applying extreme value methods in their work. In the course of organising this workshop it became apparent that the extreme value methods were of most relevance to the volcanology and space-weather teams that are located in the BGS Edinburgh office. Therefore, this workshop was moved to Edinburgh (see below). Ben Marchant interviewed four team leaders from Keyworth who had expressed an interest in attending a meeting on extreme value statistics so that their interests could be presented at the Edinburgh meeting.

These researchers were:

Vanessa Banks who was interested in modelling the occurrence of landslides. In particular she was interested in how a complete dataset of visible landslides in Yorkshire could be used to validate and improve the GeoSURE model of landslide risk and how to model reported landslides across the whole of the UK when it is known that occurrences in remote areas are under reported. These appeared to be well-contained problems that would be suitable for an MSc project and therefore a proposal will be prepared in time for the 2016-17 academic year.

Barry Rawlins who was interested in the design and analysis of efficient surveys of soil contamination. One particular concern in this work was how standard geostatistical models could be modified to ensure that the dependence structure amongst extremely polluted observations was appropriately represented. Ben Marchant included this work in his presentation to the STOR-i Centre for Doctoral Training on July 4th (see below).

Tom Dikjstra and Anna Harrison who were interested in the occurrence of landslides and residential insurance claims for subsidence, respectively. Both of these researchers were modelling these extreme events using UKPC09 realisations of future climate senarios. In both circumstances the computational requirements prevented the use of the full set of realisations and they wanted to know the most efficient ways in which the ensemble of realisations could be subsampled. This work has synergies with another SECURE Feasibility Project being conducted by LU which is concerned with whether global climate models adequately include the extreme events that lead to substantial glacial ice-melt events. When this project is completed, BGS and LU will discuss any findings that are relevant to the BGS work on landslides and subsidence.

2.4 Project Workshop

The project workshop was held at the BGS Edinburgh office on Tuesday June 21st 2016. It commenced with a seminar presentation by Emma Eastoe and Jenny Wadsworth introducing extreme value statistics. This presentation included an overview of EVT methodology and discussed its

capabilities and limitations. The results of the groundwater extremes case study were described along with details of previous applications of EVT by LU including the modelling of extreme river flow events, the downscaling of outputs of global climate models to predict the frequency of local extreme wave events and the modelling of extreme weather events that lead to the melting of glaciers in Greenland. The remainder of the workshop consisted of a series of informal meetings between Ben Marchant, Emma Eastoe and Jenny Wadsworth and various BGS research teams with the aim of initiating future collaborations between BGS and LU. In each of these meetings the BGS researchers presented problems which they believed could be addressed using EVT or more general statistical methods that were of interest to LU. A brief summary of each meeting is given below:

2.4.1 Meeting with BGS Volcanology Team (Susan Loughlin, Charlotte Vye-Brown, Samantha Engwell, Melanie Duncan and Fabio Dioguardi)

The members of the volcanology team described and presented a series of projects which require the analysis of varied and complex datasets. In some of these projects the data were still being gathered. These included attempts to use interviews and questionnaires to better understand the extent and impact of volcanic eruptions in Africa over the past 50 years. More established work included the calibration of models of the shape of volcanic debris and the implications for its dispersion, the calibration of numerical models of volcanic eruptions and the spatial analysis of volcanic deposits. It was agreed that some of these projects, particularly the spatial analysis of volcanic deposits would be suitable topics for a LU MSc project and that BPM would continue to discuss these with the Volcanology team with the aim of producing a MSc proposal by early 2017.

2.4.2 Meeting with BGS Glacier Monitoring Team (Jeremy Everest and Andrew Finlayson)

Jeremey Everest described the various sensors and monitoring equipment (e.g. seismometers, automatic weather stations, stationary GPS, boreholes, stream level sensors, streamflow sensors, and automatic cameras) that had been installed at the BGS glacier observatory at Virkisjökull, Iceland. BGS has been observing the retreat of this glacier since 2009 and the time series data of this duration are unique. Full details are available at http://www.bgs.ac.uk/research/glacierMonitoring/home.html. The group discussed how these sensor outputs could be used to produce data streams that document the retreat of the glacier and the occurrence of seismic events. It was agreed that there were many potential topics for MSc projects that would relate the ice melt and seismic events to weather variables. It was agreed that the Glacier Monitoring Team would attempt to produce an MSc project proposal for this year's course and would definitely produce further proposals by early 2017. A NERC Knowledge Exchange Call would be a potential source of money for further collaborative work beyond the completion of these MSc projects.

2.4.3 Meeting with BGS Space Weather Team (Alan Thomson, Sarah Reay and Gemma Kelly)

Alan Thomson described the problem of extreme space weather events or geomagnetic storms and described why they are included on the government's national risk register. BGS has access to high frequency (1 minute) recordings of the geomagnetic field from 28 observatories across Europe. They have previously applied EVT methodologies to the data using a standard statistical software package. This led to estimates of the largest expected storm for periods of up to 200 years (Thomson et al. 2001). It was agreed that the models in this paper could be expanded to (i) account for temporal autocorrelation in the data and clustering of extreme events (ii) to

account for the quasi-regular (period of approximately 11 years) cycles of solar activity and (iii) to consider the spatial correlation between the storms observed at different observatories. Such work is entirely consistent with the core research interests of both the BGS and LU teams and will lead to improved estimates of the likely frequency, severity and spatial extent of the storms. Two MSc projects on this subject have already commenced and some of the meeting was concerned with the practicalities of conducting these studies. The two groups remain in contact and are exploring opportunities for external funding of this work (e.g. NERC calls from the Natural Hazards and Risk research program or feasibility project funding from the SECURE network).

2.5 Meeting with BGS Seismology Group (Susanne Sargeant and Ilaria Mosca)

The BGS Seismology Group has a keen interest in the probabilistic modelling of extreme seismic events. However, rather than using EVT methods the team tend to apply methods that infer the likely magnitude and frequency of extreme seismic events from the characteristics of more normal events. EVT methods have been applied to seismic data but practitioners have been concerned that they might be wasteful of data and imprecise if the return time of the extreme is of the same order of magnitude as the duration of the available data (Knopoff Kagan, 1977). All attendees at the meeting agreed that it would be interesting to re-visit the use of EVT methods in seismology and to test whether recent theoretical developments lead to more precise estimates of the return periods. This will be the subject of a MSc project proposal for the next (2016-2017) academic year.

References

Knopoff, L., Kagan, Y., 1977. Analysis of the theory of extremes as applied to earthquake problems. Journal of Geophysical Research, 82, 5647-5657.

Thomson, A.W.P., Dawson, E.B., Reay, S.J., 2011. Quantifying extreme behaviour in geometric activity. Space Weather, 9, S10001.

2.6 Presentation at STOR-i Centre for Doctoral Training Workshop.

Ben Marchant presented BGS work on the surveying and mapping of soil contamination to the STOR-i (Statistics and Operational Research) Centre for Doctoral Training workshop on Multivariate and Spatial Extremes at Lancaster University on July 5th. The workshop had been organised by Emma Eastoe and Jenny Wadsworth. The talk focused on how standard methods of mapping soil contamination inadequately represented extreme behavior. The advice and feedback from the workshop audience has led to BGS improving their spatial models of soil properties. A journal article describing these improvements is currently being prepared. LU and BGS are preparing a PhD proposal for the Soils Training and Research Studentships Doctoral Training Centre which will address wider issues in the geostatistical prediction of soil properties.

3 Evaluation of Feasibility Project

The project offer letter suggested some specific criteria to assess the success of the feasibility project. We address these criteria below:

- **Details of who was involved in the activity (e.g. proportion of researchers, policymakers, industry people, general public, school children)**

The aim of this project was to establish links between researchers at BGS and LU. Therefore the participants were almost exclusively academic researchers (20 BGS researchers had direct contact with the primary participants in the project) with the exception of the three LU MSc Students who worked on BGS data and the PhD students who attended the presentation by Ben Marchant at the STOR-i Centre for Doctoral Training workshop. However, the various topics considered such as groundwater floods and droughts, space-weather, volcanic eruptions and the retreat of glaciers are of direct interest to policy makers and industry so we would expect the involvement of these groups as this work progresses.

- **How the funding has led, or will lead, to an application to an external funder**

  The EVT methodologies for investigating groundwater droughts have been included as elements of BGS proposals to the NERC Knowledge Exchange Program (on mapping the risk of floods) and to a consortium of representatives of the water industry (on managing urban groundwater). BGS and LU will seek opportunities to continue to colloborate on the application of EVT methods to groundwater level data. These are likely to be through NERC calls related to floods and droughts. Several other potential topics for collaboration have been identified (e.g. topics on space-weather, seismology, glacier monitoring and volcanic eruptions). The feasibility of these topics will initially be tested through MSc projects. When these projects demonstrate that there is potential for further work, BGS and LU will seek further external funding. Again, the primary source of this funding is likely to be NERC programs but the BGS experts in the different fields all have experience in obtaining funding from a variety of sources. BGS and LU are preparing a PhD proposal for submission to the Soils Training and Research Studentships Doctoral Training Centre.

- **Potential for initiating or developing long-term collaborations that will contribute to the aims of SECURE**

  This feasibility project as identified a number of research topics that are consistent with the core research interests of the LU statisticians, BGS research teams and the SECURE network. In particular, the problems on space-weather, soil contamination and seismology are likely to require innovative statistical models that will lead to more precise estimates of the risks of these hazards. The MSc projects have been identified as a sustainable mechanism to complete initial assessments of such problems. Both parties have experience in securing external funding for promising areas of research.

- **Quality, novelty and level of engagement with world-leading research**

  The EVT models of groundwater floods and droughts are more generally applicable and informative than the methodologies that are currently applied in the hydrogeological literature (see Section 4). BGS continue to apply these methods in their internally funded work, and (in collaboration with LU) will seek to publish comparisons of the EVT and standard approaches in a peer reviewed hydrology journal. The feasibility project has also directly led to BGS being able to generalise their models of soil contamination to better represent extreme behavior. A paper on this work is currently being prepared.

- **Engagement with or involvement with early career researchers**

  Six of the BGS team members who met with and and are still in contact with the primary participants in the project are early career researchers.

- **Inclusion of activities that promote public outreach and/or socio-economic impact**

The aim of the feasibility project was to promote collaborations between researchers and therefore none of the project activities were focussed on public outreach and socio-economic impact. However the topics that have been identified are of interest to the public and any new findings are likely to be communicated to the public through outreach activities such as BGS open days or blogs on the BGS website.

- **Expenditure** The total costs of the project were up to £24 665.08. The exact figure will be reported to SECURE once BGS have been invoiced by their travel supplier. The total staff costs for Emma Eastoe (16 days), Jenny Wadsworth (17 days), Ben Marchant (16 days) and John Bloomfield (4 days) were £23 599.13. Up to £1065.95 was spend on travel and subsistence for the project meetings and workshops.

4 The Case Study: Extreme Groundwater Levels

4.1 Background

The case study focused on the occurrence of groundwater floods and droughts in the UK. Droughts impact agricultural livelihoods, local and national economies and ecosystems. For example, the drought of 1975-76 is estimated to have caused £3.5bn losses to agriculture and cost the UK water industry £0.4bn (Rodda and Marsh, 2011). Floods also have devastating effects on local communities leading to loss of life, damage to infrastructure and significant financial implications with the summer floods of 2007 estimated to have cost the UK £3.2bn. Groundwater levels tend to respond more slowly than river levels to extreme storms or rainfall deficits. Therefore groundwater floods and droughts can be relatively slow to develop but they can also be more prolonged with their effects still being felt long after rainfall levels have returned to normal (Bloomfield and Marchant, 2013).

The British Geological Survey holds or has access to groundwater level (GWL) records from over 7000 boreholes in UK aquifers. Many of these records only contain a few tens of observations but GWLs have been measured regularly at some boreholes for over 100 years. This study considers two such GWL records from boreholes in the Chalk aquifer. The Chilgrove House borehole, situated in West Sussex, is believed to have the longest continuous record of GWLs in the world stretching back to 1836. Groundwater level records have been recorded at Dalton Holme (Yorkshire) since 1889. Further details about the hydrogeological setting of these boreholes can be found in MacKay et al. (2014). Figures 1 and 2 show the recorded GWLs at these sites between 1910 and 2012 alongside the corresponding precipitation and estimated potential evapotranspiration (PE) values. The PE values are based on a modified version of the Penman-Monteith equation (Monteith and Unsworth, 2008).

It is common practice in hydrological studies to standardise or normalise time series of precipitation and water levels before trying to infer relationships between them. In the surface water literature this is often achieved by fitting a parametric distribution function to the observations (McKee, 1993). However, Bloomfield and Marchant (2013) demonstrated that GWLs were less likely to conform to standard probability distribution functions and therefore used an empirical approach to transform the GWLs at a site to a normalised variable which they referred to as the Standardised Groundwater Index (SGI). Seasonal effects are removed from the GWL record by applying a different transformation to each calendar month. The monthly SGI for Chilgrove House and Dalton Holme are shown in Figures 3 and 4. A similar methodology can be used to calculate the Standardised Precipitation Index (SPI) or if the PE is subtracted from the precipitation the Standardised Effective Precipitation Index (SPEI). However, extreme events in hydrological sys-
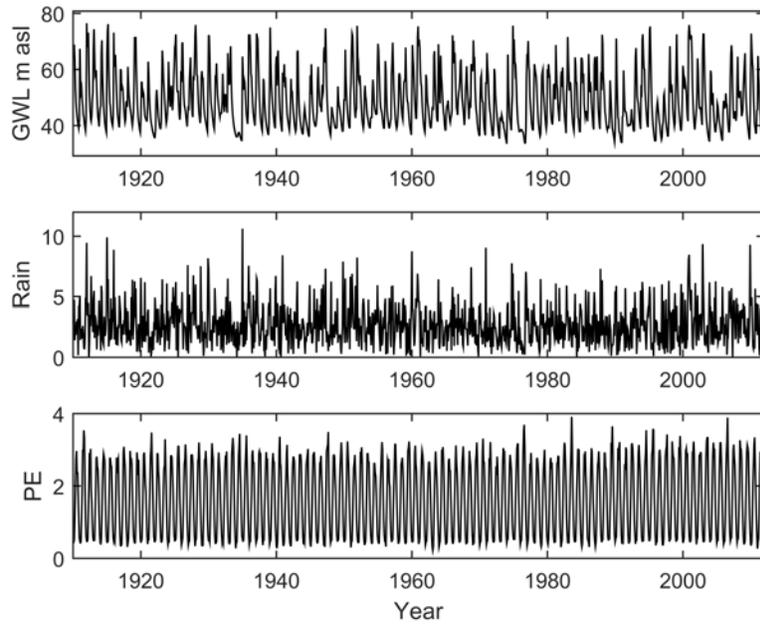
Figure 1: Observations of GWL, precipitation and PE for Chilgrove House.
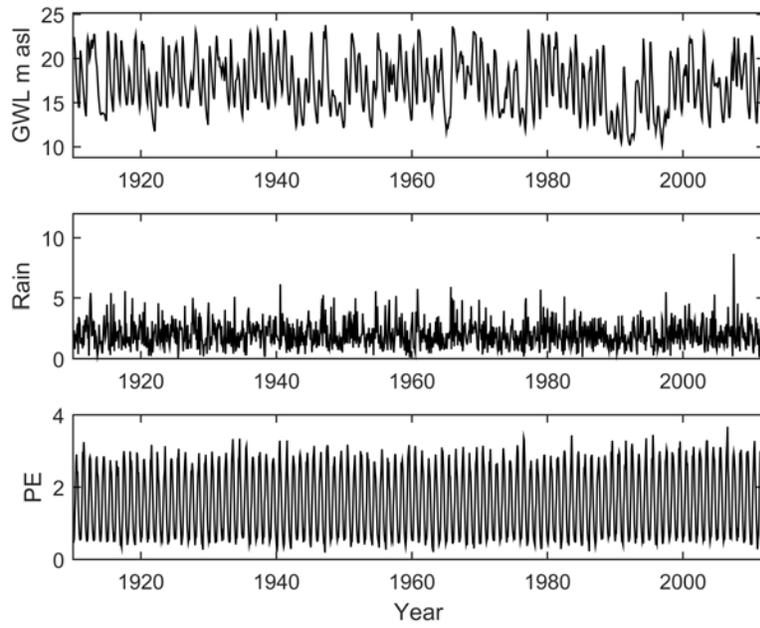
Figure 2: Observations of GWL, precipitation and PE for Dalton Holme.
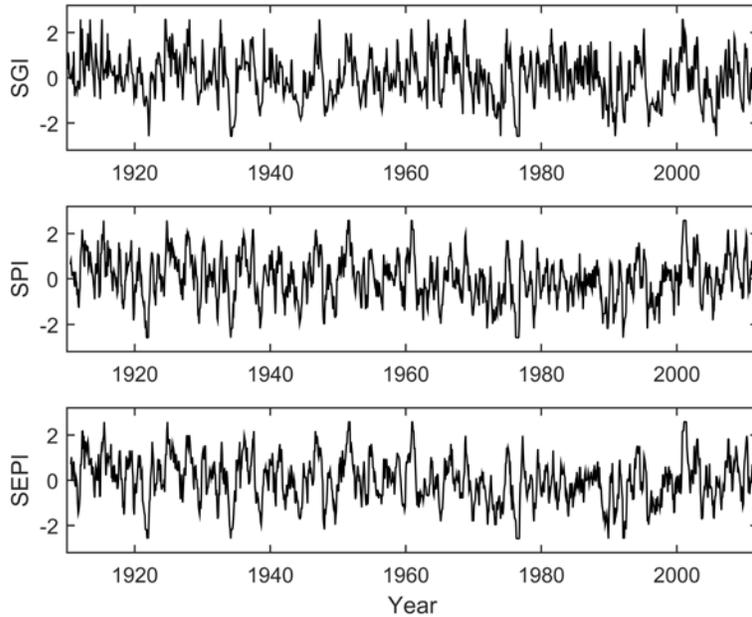
Figure 3: SGI (top), SPI (middle) and SPEI (bottom) for Chilgrove House. Accumulation period for SPI and SPEI is chosen to maximise correlation with SGI.

tems are often driven by precipitation deficits and excesses over periods longer than a single month. Therefore, prior to normalization the precipitation is summed over an accumulation period of a specified number of months.

Droughts are generally defined in terms of a deviation from the seasonal norm. This can be achieved by placing a threshold on the standardised index. For instance, McKee (1993) suggested an SPI between 0 and -1 indicated a mild drought; between -1 and -1.5 indicated a moderate drought; between -1.5 and -2 indicated a severe drought and less than -2 indicated an extreme drought. Water levels that are higher than the seasonal average can be defined similarly. However, it may be more helpful to define high groundwater levels in terms of absolute levels since these can and have been used to trigger warnings of potential groundwater flooding (Adams et al., 2010).

Studies of the variation of GWLs tend to attempt to model the entire hydrograph rather than to specifically focus of the extremes. For a number of boreholes, Bloomfield and Marchant (2013) considered the correlation between the SGI and the corresponding SPI or SPEI with various accumulation periods. They demonstrated that the accumulation period that led to the maximum correlation was an indicator of the range of the temporal correlation in the SGI. The SEPI (which accounts for PE) was slightly more strongly correlated to SGI than the SPI. The plots of the SPI and SPEI that are most strongly correlated with SGI (accumulation periods of 6 months for Chilgrove House and 10 months for Dalton Holme) are included in Figures 5 and 6.

One approach that has been used to look for specifically at GWL extremes has been proposed by Wilby et al., (2015). They assumed that the occurrence or non-occurrence of a drought (or flood) each month was the outcome of a two-state first order Markov process. The transition probabilities of either moving from drought to non-drought or non-drought to drought each month are estimated from the observational record. Then long-term sequences of the drought (or flood) indicator are simulated and various properties of the extreme event are extracted from these simulations. For example, the histrograms in Figures 7 and 8 indicate the probability distribution of the length of a

9
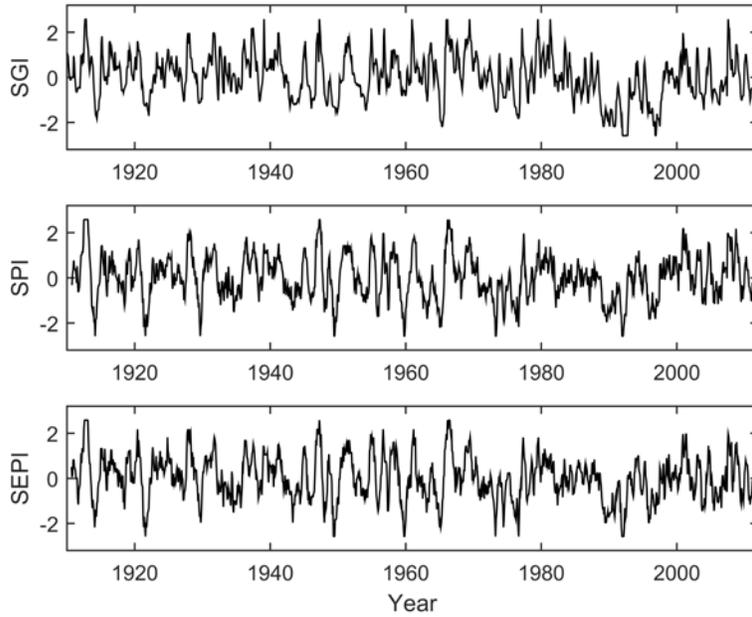
Figure 4: SGI (top), SPI (middle) and SPEI (bottom) for Dalton Holme. Accumulation period for SPI and SPEI is chosen to maximise correlation with SGI.
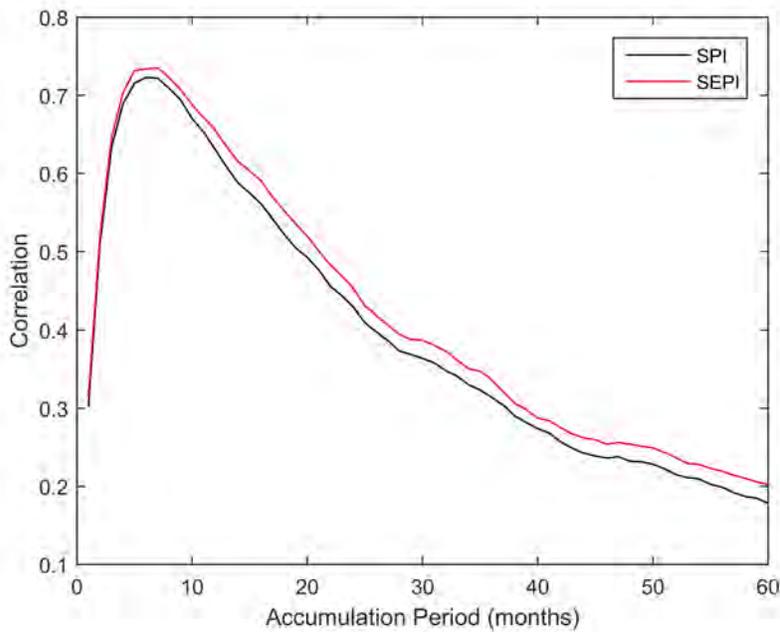
Figure 5: Correlation between SPI (black) or SPEI (red) and SGI against accumulation period for Chilgrove House.
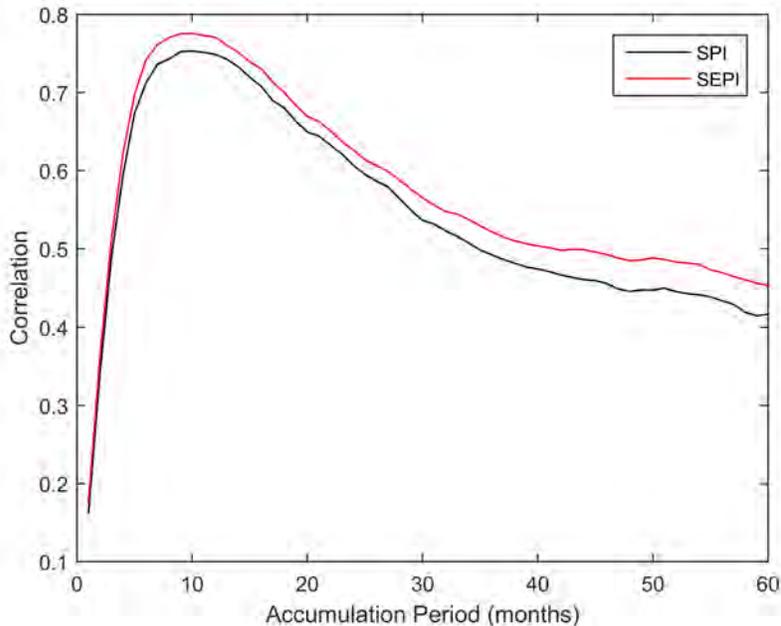
Figure 6: Correlation between SPI (black) or SPEI (red) and SGI against accumulation period for Dalton Holme.

given drought event. Properties of this distribution can also be calculated since, according to the simple transition model, the cluster length distribution will be geometric. If L is a random variable denoting cluster length and p the probability of transition from extreme to non extreme, we have

$$\Pr(L = s) = p(1 - p)^{s-1}, \quad s \in \{1, 2, 3, \ldots\}.$$

In this feasibility study we addressed two of the shortcomings with these approaches to studying GWL extremes. First, we explored how extreme value theory methodologies could be used to specifically model the occurrence and severity of an extreme GWL event rather than inferring these characteristics from models of the entire hydrograph. Second, we expanded the model of extreme dependencies to go beyond a probability of moving between extreme and non-extreme states to account for the severity of the extreme and to relax the first order Markov assumption.

4.2 Modelling the occurrence and severity of groundwater level extremes

4.2.1 Introduction

Following the common practice in the groundwater literature, both floods and droughts were modelled on a standardised scale. However, the outcomes of the flood models were reported in terms of the original GWL (in metres above sea level) since this is a more meaningful indicator of the potential impact of the flood event. Accumulated precipitation and PE were considered as covariates that potentially drive the extreme events. The accumulation periods for these two covariates were constrained to be identical since precipitation and evapotranspiration occur at or above the surface over a time scale that is considerably shorter than the flow of groundwater through the subsurface. For floods, the covariates were included on their original scale since there was no evidence for doing otherwise. For droughts, the covariates were modelled on the standardised scale.
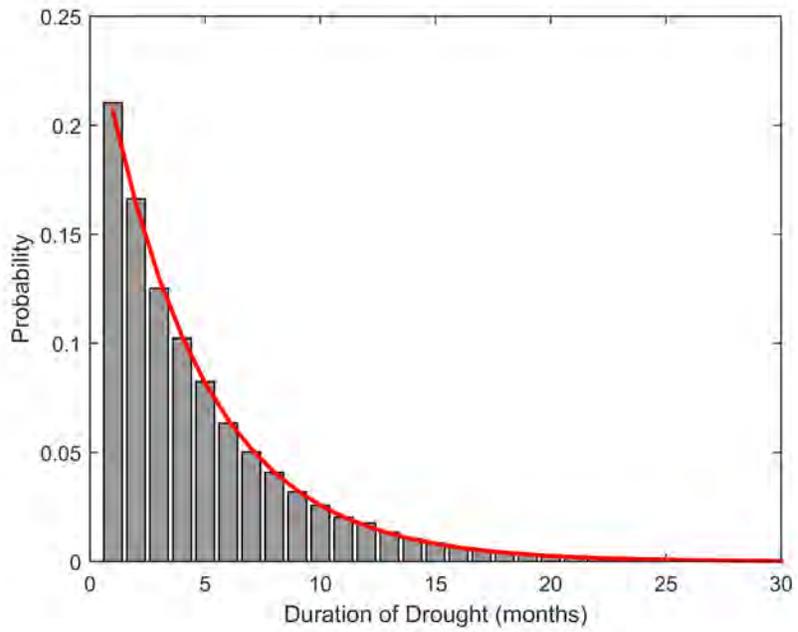
Figure 7: Simulated (bar chart) and theoretical (red) distribution of droughts of different length for Chilgrove House.
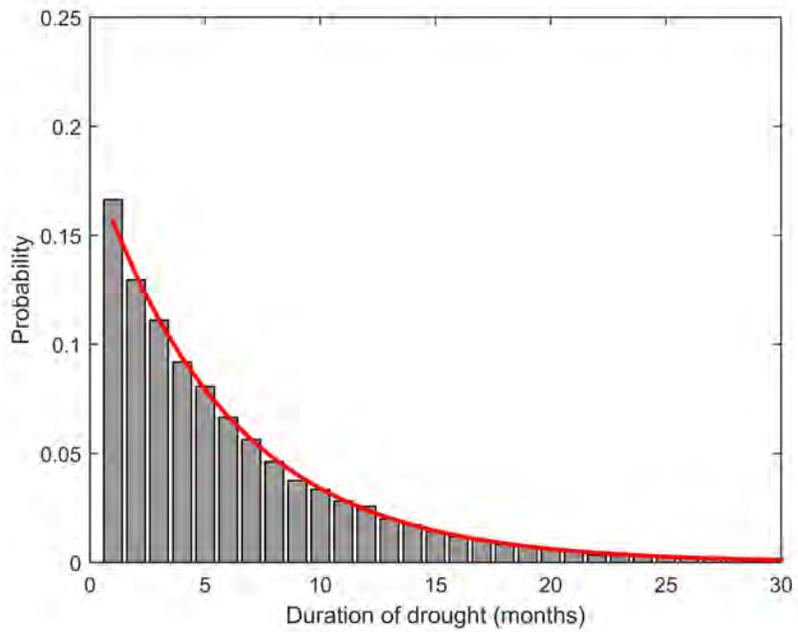
Figure 8: Simulated (bar chart) and theoretical (red) distribution of droughts of different length for Dalton Holme.

4.2.2   Models

Rate-generalised Pareto (GP) models of flood and drought events at the two boreholes were esti-mated by maximum likelihood. Likelihood-ratio tests were used to test whether the inclusion of an additional covariate significantly improved the model fit. The rate-GP (generalised Pareto) model is defined as follows. Let Y_t denote GWL (original or standardised scale as required) at time t and let $\mathbf{X}_t$ be an associated $p \times 1$ vector of covariates of accumulated precipitation and PE.

For some high threshold u, the rate of threshold exceedances

$$\phi(\mathbf{x}_t) := \Pr[Y_t > u]$$

is modelled as

$$\text{logit } \phi(\mathbf{x}_t) = \mathbf{x}_t' \boldsymbol{\phi}.$$

Here $\boldsymbol{\phi}$ is a $p \times 1$ vector of regression coefficients and the logit function is defined as

$$\text{logit}(x) = \log \left(\frac{y}{1 - y} \right).$$

For the same threshold, the size of threshold exceedances is modelled using the two-parameter GP distribution with scale parameter $\sigma > 0$ and shape parameter ξ,

$$\Pr[Y_t \leq y | Y_t > u] = 1 - \max \left\{ 0, \left[1 + \xi \left(\frac{y - u}{\sigma} \right) \right] \right\}^{-1/\xi}.$$

Similarly, for droughts we define $\phi = \Pr[Y < u]$ as the probability of being below a low threshold u, and the GP distribution function is written as

$$\Pr[Y \leq y | Y < u] = \max \left\{ 0, \left[1 + \xi \left(\frac{u - y}{\sigma} \right) \right] \right\}^{-1/\xi}.$$

Covariates can be included in the GP scale and shape parameters in a similar way to that in which they were included in the rate parameter, ϕ. In practice, there is not enough information in the data to model the shape parameter as a function of covariates, and so we assume that the shape parameter is constant and restrict ourselves to modelling

$$\log \sigma(\mathbf{x}_t) = \mathbf{x}_t' \boldsymbol{\sigma}$$

where $\boldsymbol{\sigma}$ is a $p \times 1$ vector of regression coefficients. Inference for the rate and GPD parameters can be carried out independently, by maximising the appropriate likelihood function.

Since in our case the covariates consist of accumulated rainfall and accumulated PE, we start by creating a profile likelihood for the accumulation period. This is set *a priori* to be the same for both variables, but allowed to differ for droughts and floods. Once the optimal accumulation period is chosen, the model can be fitted. The model fit can be assessed *via* plots of the fitted rate parameter, a quantile-quantile (QQ) plot of the fitted quantiles *vs.* the empirical quantiles, and a plot of simulated values super-imposed on the observed data.

The fitted rate is simply

$$\text{logit } \hat{\phi}(\mathbf{x}_t) = \mathbf{x}_t' \hat{\boldsymbol{\phi}}.$$

To obtain the QQ plot, note that if $G(Y_t|\mathbf{x}_t)$ denotes the cumulative distribution function for the GP distribution, then

$$G(Y_t|\mathbf{x}_t) \sim \text{Unif}(0, 1)$$

and so
$$- \log[1 - G(Y_t | \mathbf{x}_t)] \sim \text{Exponential}(1)$$

We can therefore plot $- \log[1 - G(y_t | \mathbf{x}_t)]$ against the theoretical quantiles from an Exponential(1) distribution to obtain a QQ plot.

Simulation relies on having a set of covariates $\mathbf{x}^*$. Given these covariates, we can then estimate the rate function $\hat{\phi}(\mathbf{x}^*)$. To decide whether or not the corresponding GWL will exceed the threshold u, simulate a uniform random variable $v \sim \text{Unif}(0, 1)$. If $v < \hat{\phi}(\mathbf{x}^*)$, then the corresponding GWL will be a threshold exceedances. This can then be simulated from the GP distribution, with parameters $\sigma(\mathbf{x}^*, \xi)$, using the probability integral transform method. Under the model defined above we have no model for non-exceedances and no way of simulating these; neither does the simulation method account for any serial dependence in the GWLs.

Having estimated the covariate model, the marginal distribution of the data set can be transformed to a common distribution to aid modelling of the lag 1 dependence structure. The standardised GWL must first be transformed to Uniform margins on $[0, 1]$ using the transformation

$$\tilde{F}(Y_t) = \begin{cases} \phi_D(x_t) \left[1 + \xi_D \left(\frac{u_D - Y_t}{\sigma_D(x_t)}\right)\right]^{-1/\xi_D} & Y_t < u_D \\ \hat{F}(Y_t) & u_D \leq y_t \leq u_F \\ 1 - \phi_F(x_t) \left[1 + \xi_F \left(\frac{Y_t - u_F}{\sigma_F(x_t)}\right)\right]^{-1/\xi_D} & Y_t > u_F \end{cases}$$

where $\hat{F}$ denotes the empirical distribution function; $\phi_D(\cdot)$, $\sigma_D(\cdot)$ and ξ_D are the fitted rate and GP scale and shape parameters for droughts; $\phi_F(\cdot)$, $\sigma_F(\cdot)$ and ξ_F are the fitted rate and GP scale and shape parameters for floods; and u_D (u_F) is the modelling threshold for droughts (floods). From Uniform margins this can be moved to any desired margin for modelling (common choices are Laplace, Fréchet or Gumbel).

4.2.3 Results

Chilgrove House: The empirical 0.1 and 0.9 quantiles of the standardised GWL data were used to define droughts and floods respectively. For the Chilgrove House drought model, the optimal precipitation and PE accumulation period is 0-8 months (where 0 months means the month in which the threshold exceedance occurred). For the corresponding floods model, the optimal precipitation and PE accumulation period is 0-2 months. Plots of the profile likelihoods for these periods are given in Figure 9. These different accumulation periods are consistent with the observation of Eltahir and Yeh (1999) that groundwater drought episodes tend to be more prolonged than floods. The corresponding exceedance probabilities are plotted in Figure 10.

Examples of data simulated from both drought and flood models are given in Figure 11. In both cases the covariates used in simulation were the observed covariates. Simulated data are given on the standardised scale (droughts) and the original scale (floods). For the floods, because the model was fitted on the standardised scale, the simulated values had to be transformed back to the original scale using observed monthly means and standard deviations of the GWL.

Figure 12 shows the dependence between successive monthly observations once the data have been transformed to Laplace margins. The first figure shows the data as a time series, the second is a scatter plot of the observations at lag t against the observations at lag $t-1$. This transformation appears to lead to a more symmetric behaviour in the dependence structure than that observed by transforming without accounting for covariates directly.
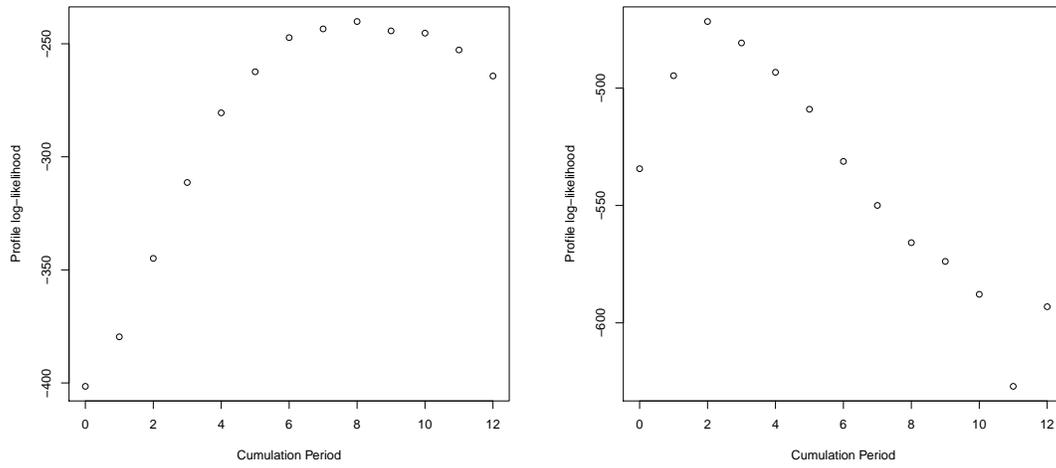
Figure 9: Profile likelihoods for accumulation period (months) for threshold exceedance models: droughts (left) and floods (right).
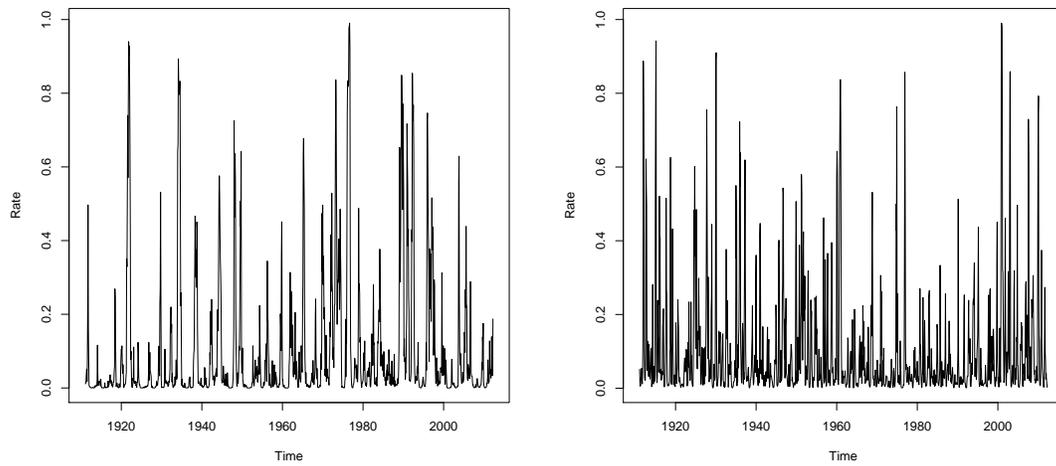
Figure 10: Estimated probability of a threshold exceedance for droughts (left) and floods (right).

15
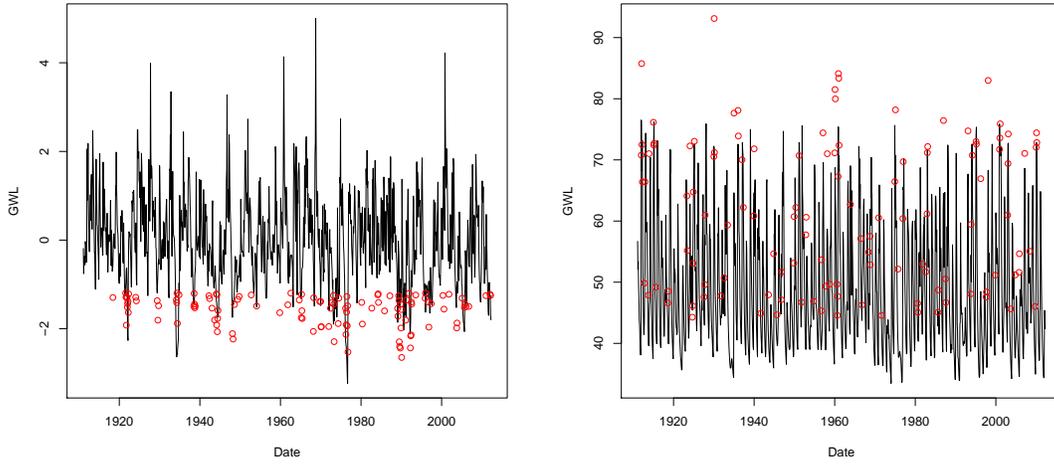
Figure 11: Simulated data (red circles) for droughts (left) and floods (right). Black lines show the observed GWL on the standardised (drought) and original (flood) scales.
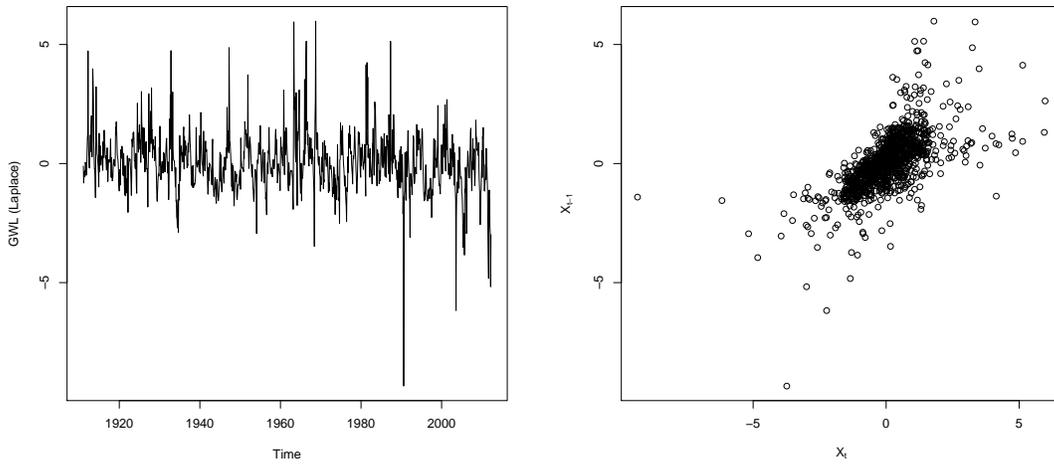
Figure 12: GWL transformed to common Laplace margins: time series (right-hand side) and lag 1 dependence (left-hand side).
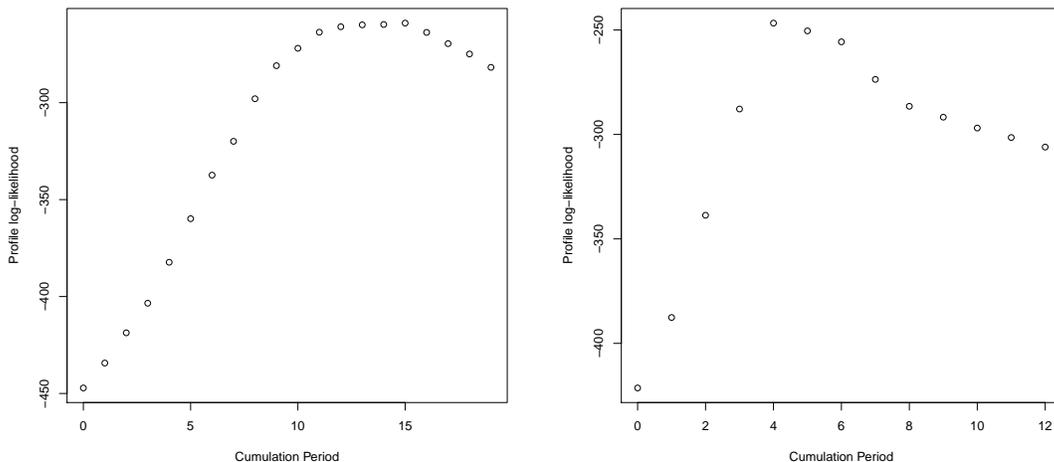
Figure 13: Dalton Holme: profile likelihoods for cumulation period for threshold exceedance models: droughts (left) and floods (right).

Dalton Holme: The analyses were repeated for Dalton Holme. Here the optimal precipitation and PE accumulation periods were 0-15 months (droughts) and 0-4 months (floods). It should be noted from the plots in Figure 13 that the profile likelihood for the accumulation period for droughts is fairly flat, particularly between 11 and 16 months. Plots of the rate parameters and simulated values for both droughts and floods are given in Figure 14 and 15 for probabilities and simulations respectively.

4.3 Exploratory extremal dependence analysis in groundwater time series

Exploratory extremal dependence analyses were carried out to gather an initial assessment on the nature of the extremal dependence in both tails of the GWL time series, $\{Y_t\}$. This focussed on use of two summary measures, applied to pairs of consecutive observations (Y_t, Y_{t+1}).

The first summary is the coefficient of tail dependence, (Ledford and Tawn, 1996). Suppose that (X_1, X_2) are two random variables with standard exponential marginal distributions (i.e. $\Pr(X_1 > x) = \Pr(X_2 > x) = e^{-x}$, $x > 0$); this is achieved approximately in practice via marginal transformations. Then the coefficient of tail dependence is defined as the parameter $\eta \in (0, 1]$, in the model

$$\Pr(X_1 > x, X_2 > x) = \Pr(\min(X_1, X_2) > x) = Ce^{-x/\eta}, \quad C > 0,$$

which is assumed to hold for $x > x^*$, where x^* is a high quantile (often the 90% or 95% quantile). Larger values of η correspond to stronger extremal dependence, with $\eta < 0.5$, $\eta \approx 0.5$, $\eta > 0.5$ corresponding to negative extremal association, near extremal independence, and positive extremal association, respectively. The value $\eta = 1$ plays a special role, as this suggests that the data are *asymptotically dependent*, which roughly means that there will always be a positive probability of observing joint extremes of (Y_t, Y_{t+1}), whatever the level of the extremes.

A complementary measure of extremal dependence is the measure $\chi(u)$, defined for two variables (U, V) with standard uniform marginal distributions (i.e. $\Pr(U > u) = \Pr(V > u) = 1 - u$, $0 <$

17
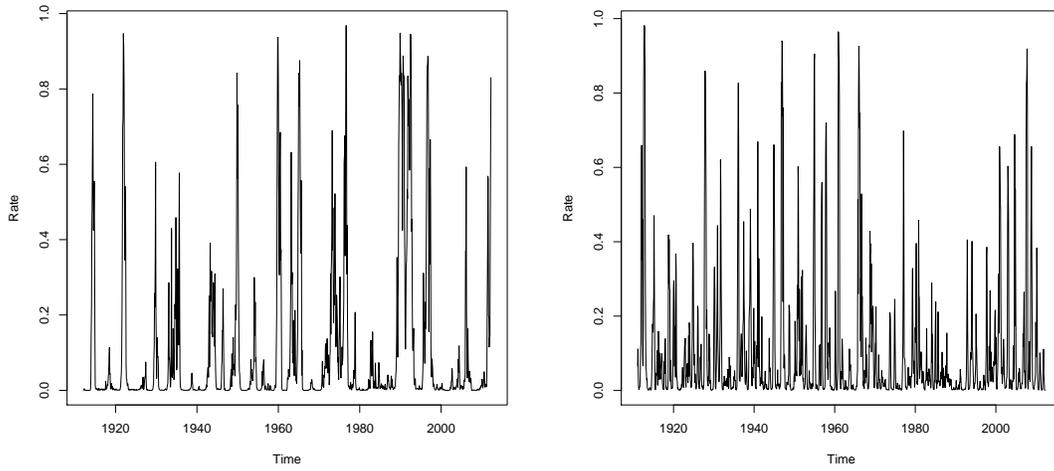
Figure 14: Dalton Holme: Estimated probability of a threshold exceedance for droughts (left) and floods (right).
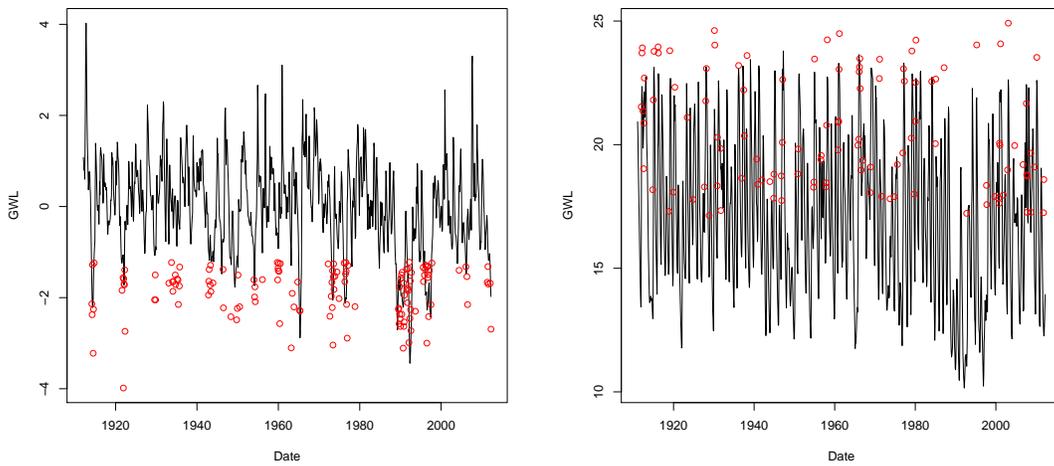
Figure 15: Dalton Holme: Simulated data (red circles) for droughts (left) and floods (right). Black lines show the observed GWL on the standardised (drought) and original (flood) scales.

$u < 1$) as

$$\chi(u) = \frac{\Pr(U > u, V > u)}{1 - u}.$$

Of particular interest is the behaviour of $\chi(u)$ as u approaches 1, as if this remains positive, this is indicative of asymptotic dependence ($\eta = 1$), whilst if it tends to zero, this is indicative of asymptotic independence ($\eta < 1$). Canonically both these summary measures focus on the upper tail, thus when we wanted to estimate them in the context of the lower tail, we transformed the lower tail to be the upper tail, which is achieved by applying a monotonically decreasing marginal transformation.

In this initial stage of exploratory extremal dependence analysis, marginal transformation took account of the value of the PE covariate, as exploratory plots revealed this to have a notable impact on the lag 1 dependence structure. Tentative conclusions were that the lag 1 dependence was stronger in the lower tail than the upper tail and that the value of PE had some effect on the level of dependence; in particular lower tail dependence was stronger for larger values of PE. These findings are consistent with our understanding of the groundwater system. Droughts are less likely to break because for GW to recharge and lead to GWL rise the soil needs to wet up and wetted pathways from the surface to the water table need to be developed first before rainfall can become effective in leading to GWL rise. Whereas under high GWL stands as soon as rainfall stops the system will drain and GWLs will fall.

During the exploratory extremal dependence analysis, marginal transformations were made according to one of four PE groupings, which somewhat represent seasonality in the data. An alternative way to standardise the data to remove seasonal heterogeneity is to assume that observations from a particular calendar month come from a common distribution. This results in a time series of Standardised Groundwater Index, as described in Section 4.1. To go forward and model the dependence, we focussed on standardised GWLs and on modelling the lower tail dependence.

4.4 Modelling Lower Tail Dependence of Standardised GWLs

Full statistical models are required to determine the expected duration of an extreme event. The approach described below is based on the model of Heffernan and Tawn (2004) which can account for dependencies in both forwards and backwards directions and hence also accomodate asymmetry observed in the data.

4.4.1 Methods

Marginal Transformations The GWL series at Chilgrove House and Dalton Holme were standardised seasonally to uniformity by using the following transform within each month $j \in \{1, \ldots, 12\}$ of the data:

$$\tilde{F}_j(Y_t) = \begin{cases} \hat{F}_j(Y_t) & Y_t > u^* \\ \hat{H}(Y_t) & Y_t < u^* \end{cases}$$

where $\hat{F}_j$ denotes the empirical distribution function of the observations in month j, and $\hat{H}$ is the generalised Pareto distribution function fitted to the lower tails pooled across all months (this was reasonably well supported by diagnostic plots; furthermore fits within individual months of the GP distribution were quite problematic, due to very short tails in some months).

The resulting time series can then be transformed to any scale desired; we denote the marginally transformed series by $\{X_t\}$. The left panel of Figure 16 shows the time series on a standard normal scale (i.e., each month has been transformed to standard normal), whilst the right panel of Figure 16 shows (X_t, X_{t+1}) on this transformed scale.

19
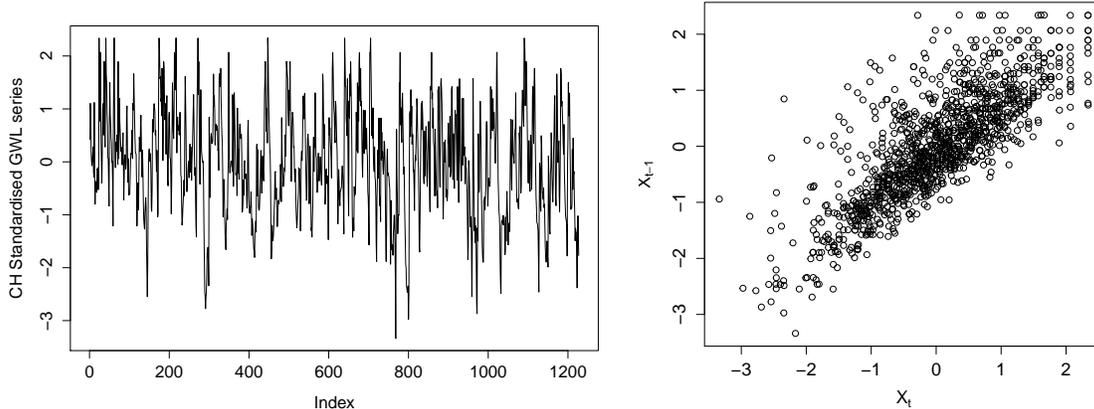
Figure 16: Left: standardised GWL time series; Right: lag 1 dependence

To analyse the extremal dependence, we transformed to standard exponential margins, with the lower tail transformed to be the upper tail; see Figure 17. The explicit goal here was to model the excursions of the time series over a high threshold (corresponding to low values of the original series) using models motivated by extreme value theory. Specifically we adopted a version of the model of Heffernan and Tawn (2004), recently used in a time series context by Winter and Tawn (2016). We began with a simple assumption that the time series behaves as a first order Markov chain in the extremes (i.e. X_t is independent of $X_{t-k}, k \geq 2$ given X_{t-1}). This simple analysis helps to illustrate the theory, but appeared to be too restrictive, thus ultimately a slightly more sophisticated model will be described.
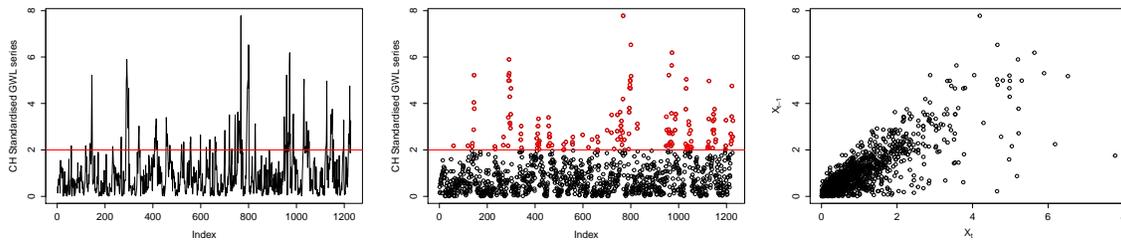
Figure 17: Left: standardised groundwater time series in exponential margins, with upper tail as the lower tail; Centre: same as left, but with red points highlighting "clusters"; Right: Lag 1 dependence in exponential margins.

First order conditional extremes model Heffernan and Tawn (2004) show that for random variables in exponential-tailed margins, a broadly valid assumption is

$$\Pr\left(\frac{X_{t+1} - \alpha X_t}{X_t^\beta} \leq x, X_t - u > y \mid X_t > u\right) \to G(x)e^{-y}, \quad u \to \infty,$$

i.e. large values of X_t become independent of $\{X_{t+1} - \alpha X_t\}/X_t^\beta$, which follow some unknown non-degenerate distribution G. When the variables have Laplace margins, the parameters $\alpha \in [-1, 1], \beta \in (-\infty, 1)$. If instead exponential margins are used and dependence is non-negative,

20

then $\alpha \in [0,1]$; this constraint was implemented in the fits described below. In either case, the parameters α, β may be estimated by maximum likelihood under a false working assumtion that G is Gaussian; confidence intervals may be obtained through bootstrap procedures, but we do not consider this here. Using the maximum likelihood estimates (MLEs) of α, β, the residuals $Z = \{X_{t+1} - \hat{\alpha} X_t\} / X_t^{\hat{\beta}}$ can be calcuated, and their empirical distribution used to approximate G.

The main reason for fitting a model is to infer characteristics of the extremes: here this is achieved through simulation from the fitted model. To simulate new values of $(X_t, X_{t+1}) | X_t > u$, denoted $(\tilde{X}_t, \tilde{X}_{t+1})$, the following algorithm was used:

Algorithm 1

1. Draw $\tilde{X}_t \sim \text{Exp}(1) + u$

2. Draw a value Z_i from the empirical distribution of residuals (with replacement)

3. Set $\tilde{X}_{t+1} = \hat{\alpha} \tilde{X}_t + \tilde{X}_t^{\hat{\beta}} Z_i$
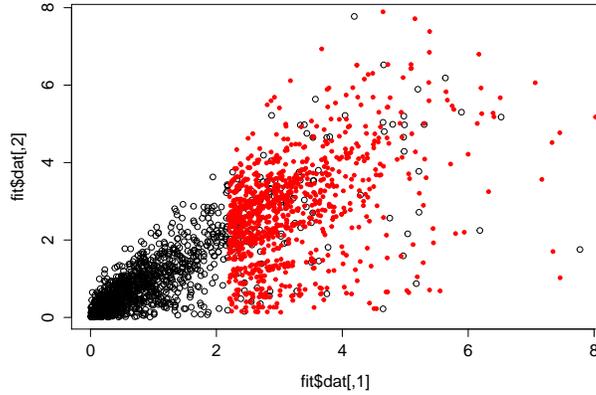
Figure 18: Lag 1 dependence at Chilgrove House in standard exponential margins (black points) with new values simulated (red points)

Algorithm 1 is illustrated in Figure 18. Under the assumption that the time series is first order Markov, a simulation scheme like this can be iterated to produce a chain of values $(\tilde{X}_t, \tilde{X}_{t+1}, \ldots, \tilde{X}_{t+l})$ larger than a high threshold u. Specifically,

Algorithm 2

1. Draw $\tilde{X}_t \sim \text{Exp}(1) + u$. Set $k = 0$.

2. While $\tilde{X}_{t+k} > u$:

   - Draw a value Z_i from the empirical distribution of residuals (with replacement)
   - $\tilde{X}_{t+k} = \hat{\alpha} \tilde{X}_{t-k} + \tilde{X}_{t-k}^{\hat{\beta}} Z_i$
   - Set $k \to k+1$

   Else stop

21

3. Return the tail chain $(\tilde{X}_t, \ldots, \tilde{X}_{t+l})$.

Algorithm 2 describes how to simulate a tail chain in a forwards direction, but this procedure could also be done in the other direction, i.e. simulating backwards. This approach and its application to Chilgrove House is described in Section 4.4.3.

To determine if this model is adequate for the tails of the transformed GWL time series, we used repeated simulation to assess its ability to capture the characteristics of the observed data. Specifically we focussed on three characteristics: (i) cluster length; (ii) cluster maximum; (iii) average value of exceedances. Clusters here were defined as groups of uninterrupted threshold exceedances. Alternatively one could use the Runs method (Smith and Weissman, 1994), whereby threshold exceedances followed by a run of at least m consecutive non-exceedances are deemed to be in separate clusters, else they are deemed to be in the same cluster.

4.4.2 Results

Comparison of cluster characteristics: Chilgrove House   The first-order Markov conditional extremes model was used to create 1000 simulated clusters at various thresholds. Figures 19-21 display histograms of the summaries (i), (ii) and (iii), whilst Tables 1-3 gives the average value of these summaries. In both the figures and tables the first row corresponds to the observed data, the second to the simulated data from the first-order Markov conditional extremes model. The third row corresponds to the methodology to be described in Section 4.4.3. (Note the differing scales on the histograms.) Numbers in parentheses are the standard deviations of the observed means.

Threshold $u = 2$ Number of observed clusters: 47.

	Mean cluster length	Mean cluster maximum	Mean cluster mean
Observed	3.40 (0.45)	3.31 (0.20)	2.75 (0.09)
Simulated	4.4	3.83	3.13
Simulated (second order)	4.62	3.58	3.07

Table 1: Observed and simulated cluster characteristics with threshold $u = 2$
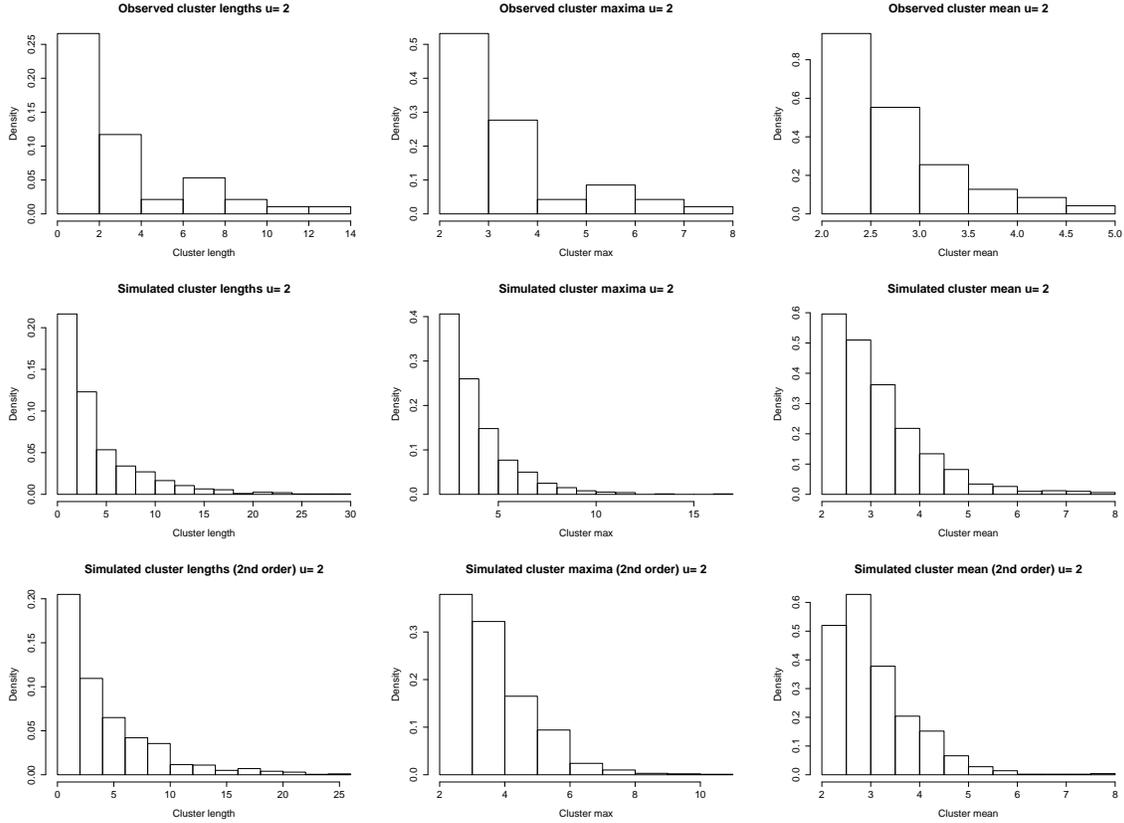
Figure 19: Observed and simulated cluster characteristics with threshold $u = 2$

Threshold $u = 2.25$ Number of observed clusters: 44.

23

	Mean cluster length	Mean cluster maximum	Mean cluster mean
Observed	2.66 (0.38)	3.44 (0.20)	3.00 (0.10)
Simulated	3.47	4.03	4.40
Simulated (second order)	3.91	3.92	3.36

Table 2: Observed and simulated cluster characteristics with threshold $u = 2.25$
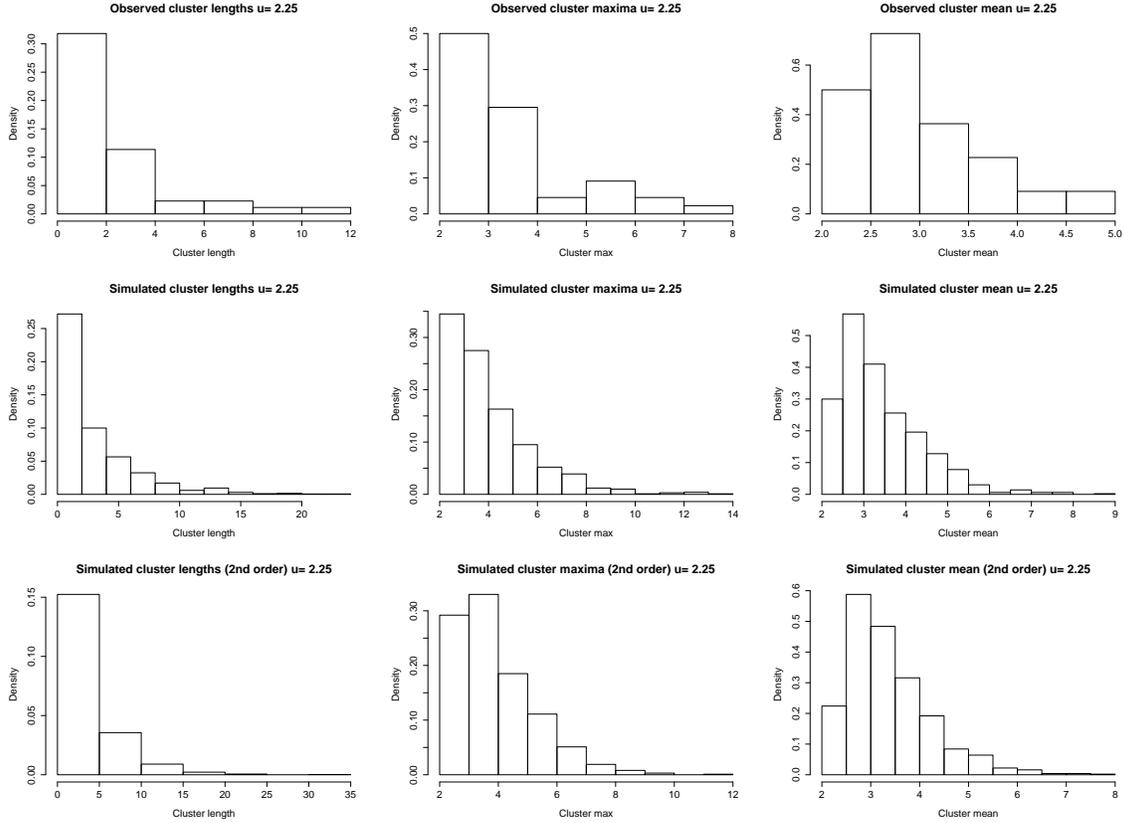
Figure 20: Observed and simulated cluster characteristics with threshold $u = 2.25$

Threshold $u = 2.5$ Number of observed clusters: 34.

	Mean cluster length	Mean cluster maximum	Mean cluster mean
Observed	2.76 (0.44)	3.75 (0.23)	3.29 (0.12)
Simulated	3.56	4.35	3.69
Simulated (second order)	3.53	4.16	3.63

Table 3: Observed and simulated cluster characteristics with threshold $u = 2.5$
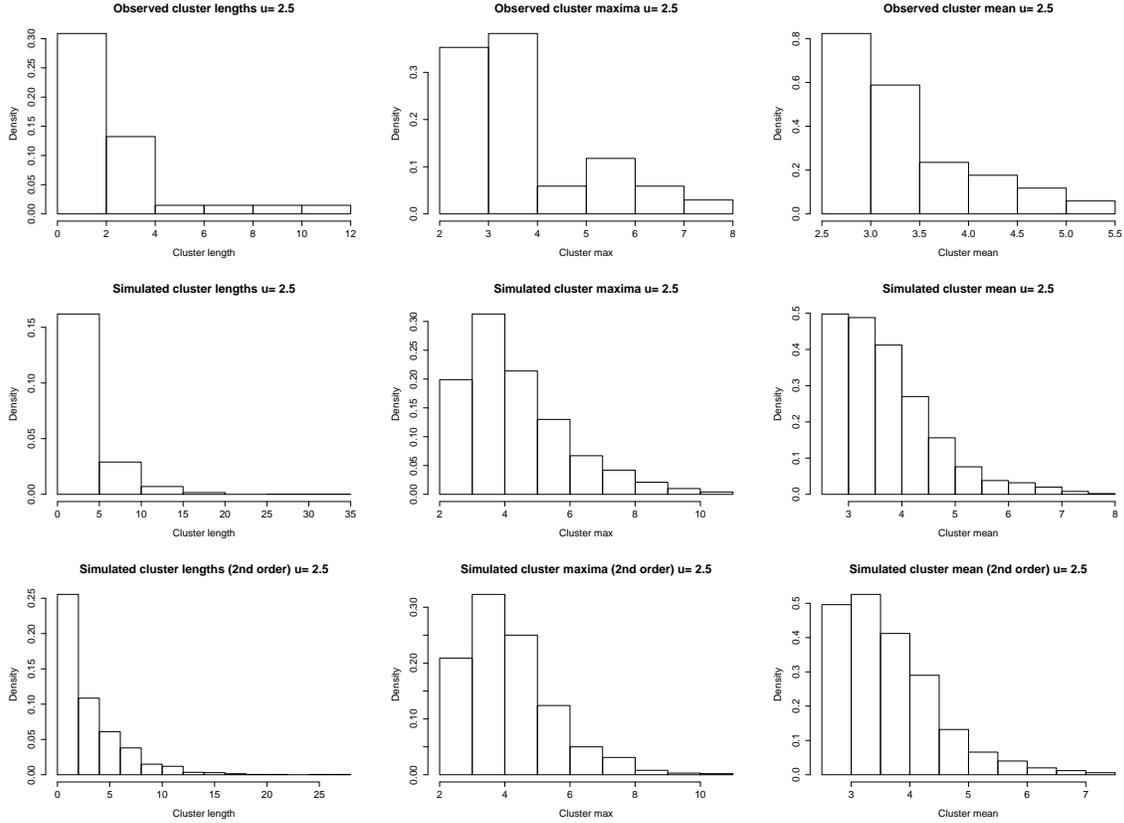
Figure 21: Observed and simulated cluster characteristics with threshold $u = 2.5$

For each threshold, the clusters simulated using the first-order Markov chain appear to be too long, with the simulated cluster maxima and means are also systematically too large. This dataset was examined again using backwards simulation, described in Section 4.4.3.

Comparison of cluster characteristics: Dalton Holme   The previous analysis was repeated for Dalton Holme; the standardised data (X_t, X_{t+1}), are plotted in Figure 22, with the lower tail as the upper tail. Figures 23-25 and Tables 4-6 summarise the simulations as above.
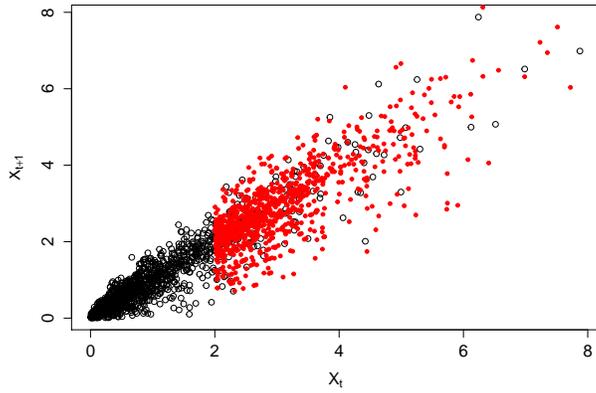
Figure 22: Lag 1 dependence for Dalton Holme in standard exponential margins (black points) with new values simulated (red points)

Threshold $u = 2$ Number of observed clusters: 28.
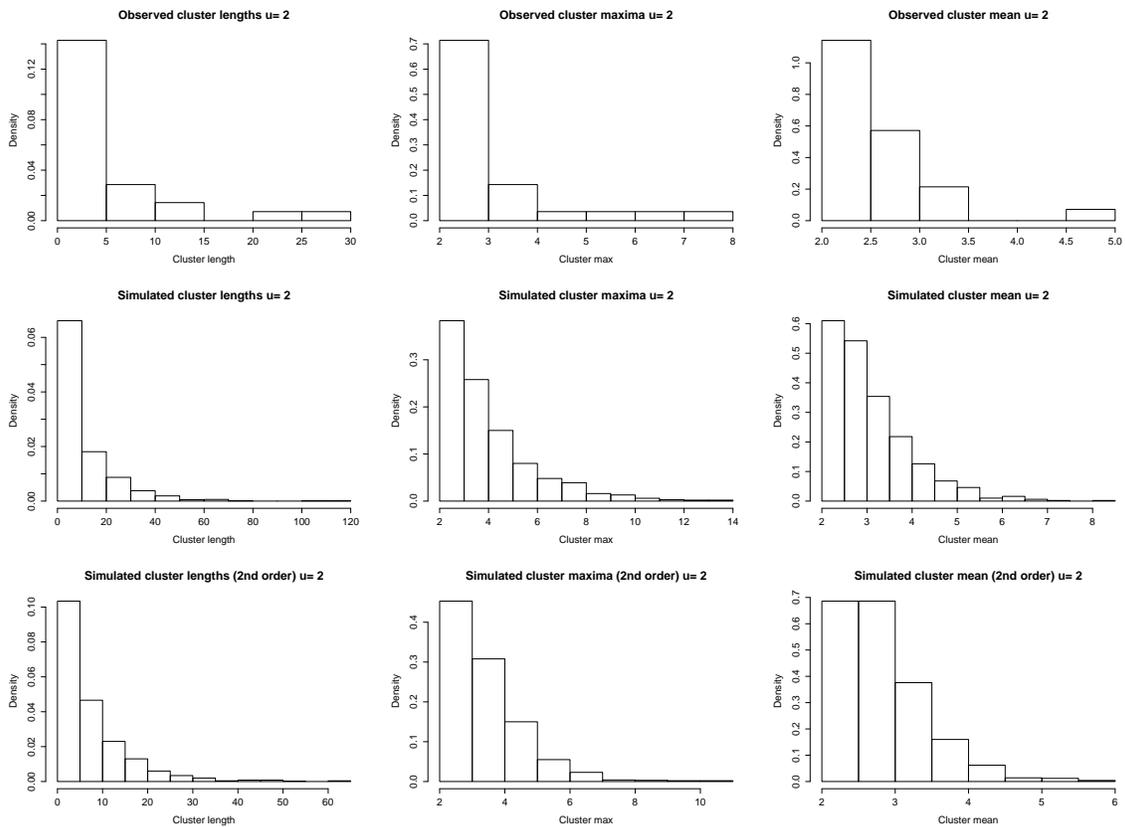
Figure 23: Observed and simulated cluster characteristics with threshold $u = 2$

Threshold $u = 2.25$ Number of observed clusters: 19.

	Mean cluster length	Mean cluster maximum	Mean cluster mean
Observed	5.71 (1.26)	3.00 (0.26)	2.54 (0.11)
Simulated	10.6	3.95	3.07
Simulated (second order)	7.77	3.40	2.82

Table 4: Observed and simulated cluster characteristics with threshold $u = 2$

	Mean cluster length	Mean cluster maximum	Mean cluster mean
Observed	6.16 (1.64)	3.43 (0.34)	2.85 (0.14)
Simulated	9.30	4.17	3.31
Simulated (second order)	7.11	3.70	3.10

Table 5: Observed and simulated cluster characteristics with threshold $u = 2.25$
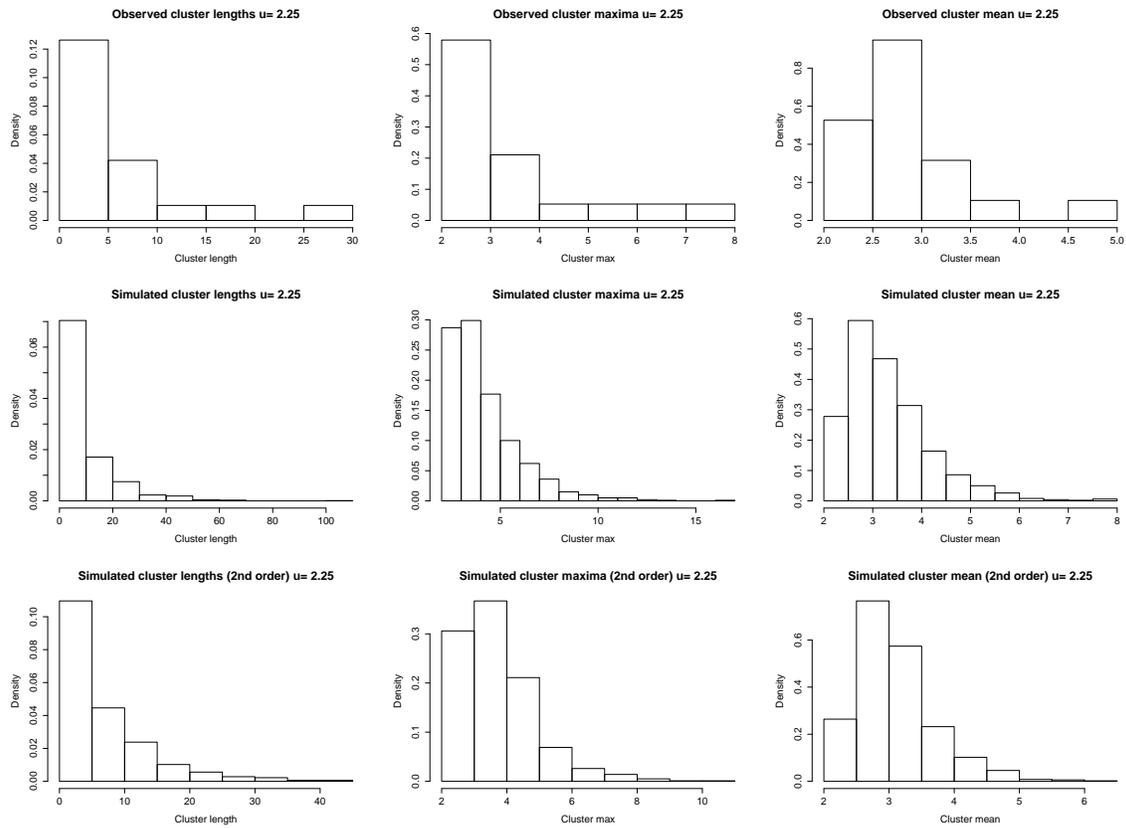
Figure 24: Observed and simulated cluster characteristics with threshold $u = 2.25$

Threshold $u = 2.5$ Number of observed clusters: 17.

	Mean cluster length	Mean cluster maximum	Mean cluster mean
Observed	5.59 (1.60)	3.63 (0.36)	3.07 (0.16)
Simulated	8.08	4.35	3.51
Simulated (second order)	5.92	3.88	3.34

Table 6: Observed and simulated cluster characteristics with threshold $u = 2.5$
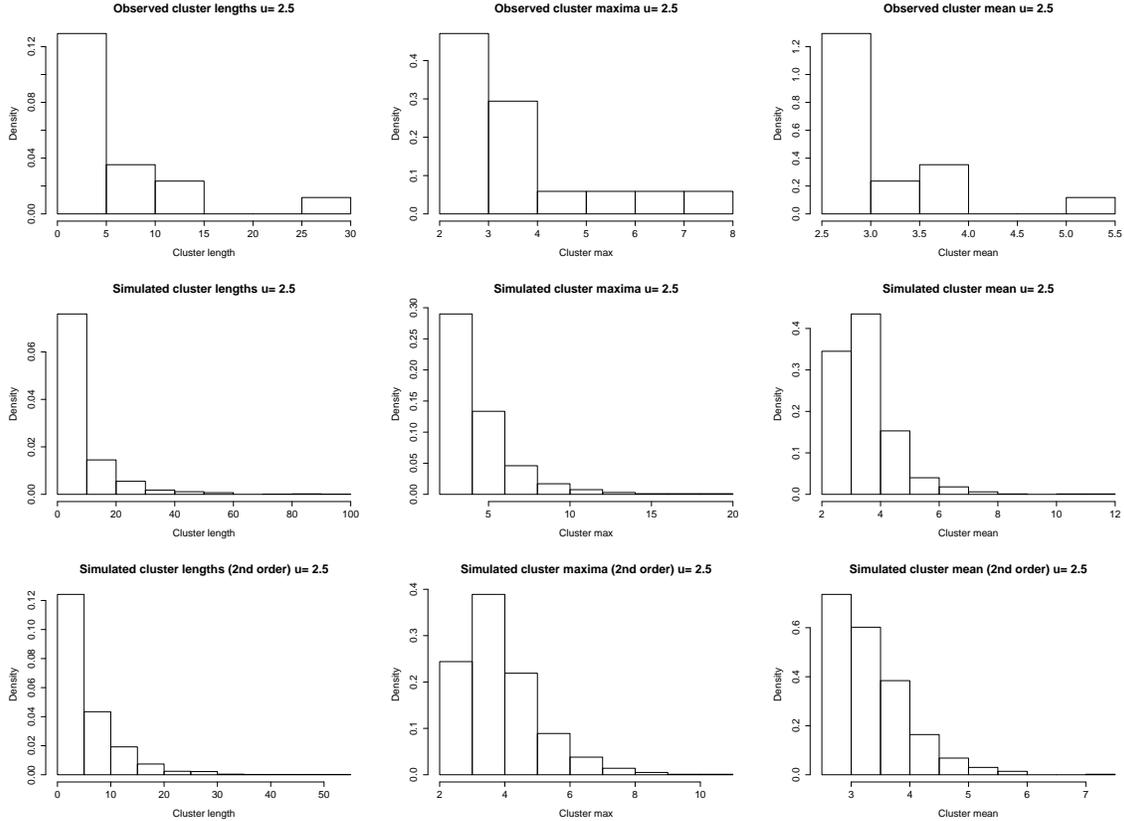
Figure 25: Observed and simulated cluster characteristics with threshold $u = 2.5$

As with Chilgrove House, the first order model overestimated cluster lengths, maxima and means. We now describe a proposal to improve the model.

4.4.3 Extensions

Second order conditional extremes model As the first order model did not capture the cluster characteristics well, we proposed a modification that allows for some consideration of the previous two values of the chain. The motivation for this came from exploratory analysis of the partial autocorrelation function (PACF), calculated from the data on the Gaussian scale, which for both the Chilgrove House and Dalton Holme shows negative dependence at lag 2.
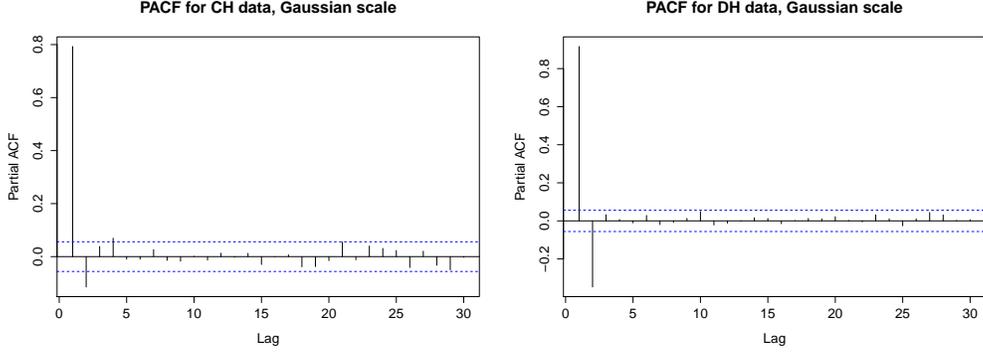
Figure 26: PACFs for Chilgrove House and Dalton Holme

Although the PACFs include dependence throughout the body as well as the tails, Figure 26 hints at the possibility of negative dependence between $(X_t, X_{t+2})|X_{t+1}$ (although positive dependence unconditionally). One way to assess whether there is indeed negative dependence in the extremes is to perform both conditional extremes fits both one step forwards and one step backwards and examine association in the residuals. Figure 27 displays these for Dalton Holme, confirming negative association (Kendall / Spearman / Pearson correlations: $-0.23/-0.34/-0.33$).
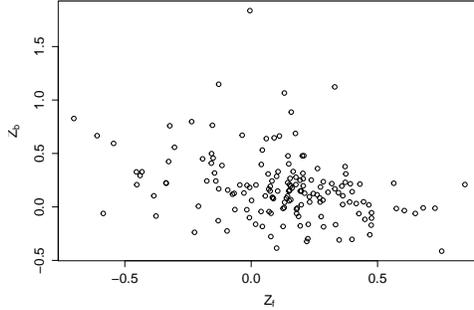
Figure 27: DH: One step forwards and backwards residuals

In order to account for this, we proposed the following novel algorithm:

Algorithm 3

1. Use the one step forwards conditional fit (with parameters (α_F, β_F)) and calculate residuals $Z_F = \{X_{t+1} - \hat{\alpha}_F X_t\}/X_t^{\hat{\beta}_F}$

2. Use the one step backwards conditional fit (with parameters (α_B, β_B)) and calculate residuals $Z_B = \{X_t - \hat{\alpha}_B X_{t+1}\}/X_{t+1}^{\hat{\beta}_B}$

3. Concatenate forwards–backwards residuals (Z_F, Z_B)

4. Draw $\tilde{X}_t \sim \text{Exp}(1) + u$

5. Draw a value Z_F from the empirical distribution of forwards residuals (with replacement)

6. Set $\tilde{X}_{t+1} = \hat{\alpha}_F \tilde{X}_t + \tilde{X}_t^{\hat{\beta}_F} Z_F$

7. Use (α_B, β_B) to find the corresponding backwards residual $\tilde{Z}_B = \{\tilde{X}_t - \hat{\alpha}_B \tilde{X}_{t+1}\}/\tilde{X}_{t+1}^{\hat{\beta}_B}$

8. Sample the next residual $Z_F|\tilde{Z}_B$ (see below for a possible method)

9. Proceed as in Algorithm 2, but at each step following items 6–8 to ensure conditional sampling of Z_F.

To sample $Z_F|\tilde{Z}_B$, we adopted the simplest method of using "nearest neighbours" in the Z_B space. That is, for a chosen value of m, we selected the m nearest values of Z_B to $\tilde{Z}_B$, and sampled the next Z_F only from those associated to the relevant Z_B values. This is illustrated in Figure 28. A more sophisticated approach would be to use kernel density estimation to produce weights for the conditional sampling, but this was not considered further here. Following this approach, if we simulate a large value $\tilde{X}_{t+1}$, then this leads to a larger value of $\tilde{Z}_B$, which in turn means we sample smaller values of Z_F for the next step forwards, resulting in shorter chains.
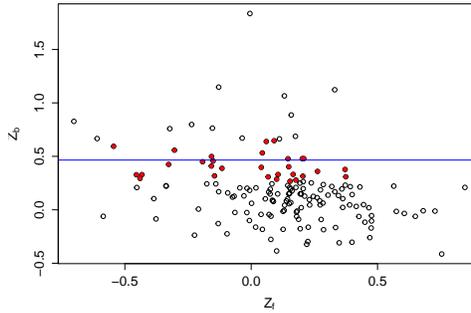
Figure 28: DH: One step forwards and backwards residuals, with conditional sampling illustrated for $m = 30$. Blue line indicates the value of $\tilde{Z}_B$, red dots show the values of Z_F that can be sampled for the next value.

The results of applying this second order algorithm are given in Section 4.4.2 for Chilgrove House and Dalton Holme, in the third rows of the tables and figures (with $m = 20$ nearest neighbours used). For Chilgrove House, the effect is small, which perhaps reflects the fact that the negative correlation in the forward–backwards residuals is not as strong (Kendall / Spearman / Pearson correlations at $u = 2$: $-0.095/-0.12/-0.10$) as that of Dalton Holme, where reasonable improvements are noted.

Backward simulation for Chilgrove House Neither the first nor second order simulation schemes for Chilgrove House appeared to satisfactorily capture the empirical characteristics of the sample. Noting that these data have a higher degree of asymmetry than the Dalton Holme data, we also experimented with backwards simulation. Figures 30-32 and Tables 7-9 are as in section 4.4.2, except the second row now corresponds to first order backwards simulation. Figure 29 displays the lag 1 values, beginning with $X_{t+1} > u$.

30

	Mean cluster length	Mean cluster maximum	Mean cluster mean
Observed	3.40 (0.45)	3.31 (0.20)	2.75 (0.09)
Simulated (backwards)	3.96	3.47	2.84

Table 7: Observed and backwards simulated cluster characteristics with threshold $u = 2$
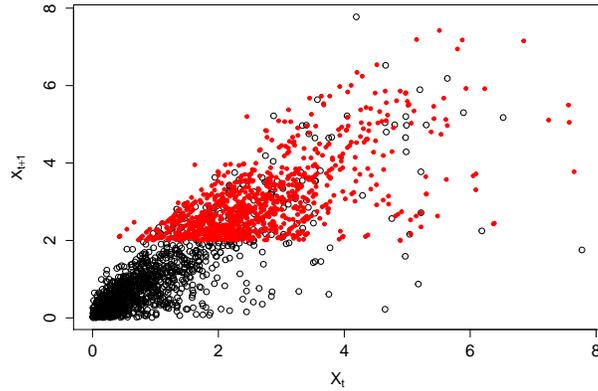
Figure 29: Lag 1 dependence for Chilgrove House in exponential margins (black points) with new values simulated (red points)
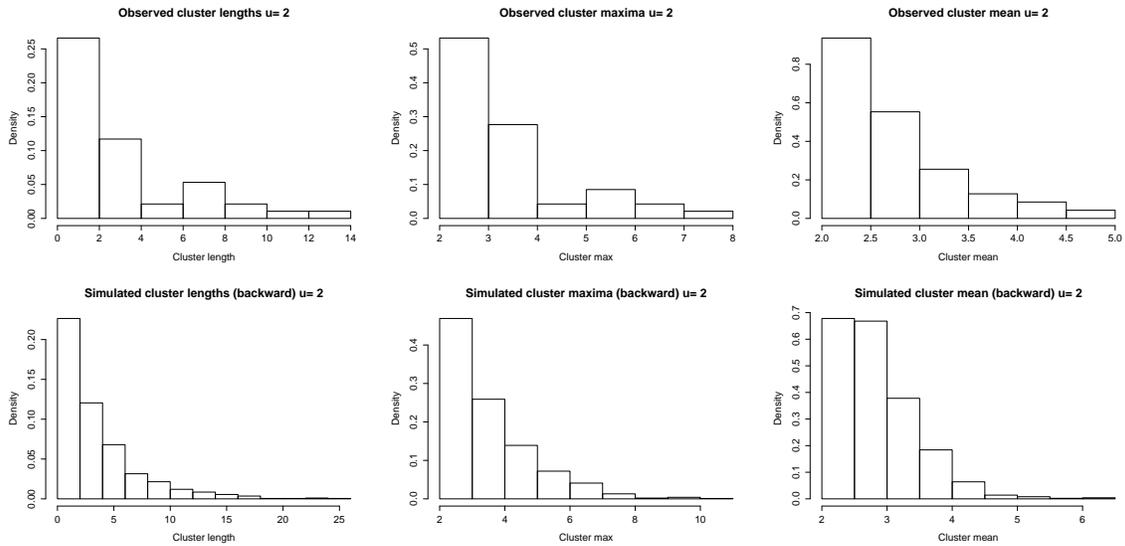
Threshold $u = 2$

Figure 30: Observed and backwards simulated cluster characteristics with threshold $u = 2$

Threshold $u = 2.25$

	Mean cluster length	Mean cluster maximum	Mean cluster mean
Observed	2.66 (0.38)	3.44 (0.20)	3.00 (0.10)
Simulated (backward)	2.96	3.46	3.02

Table 8: Observed and backwards simulated cluster characteristics with threshold $u = 2.25$
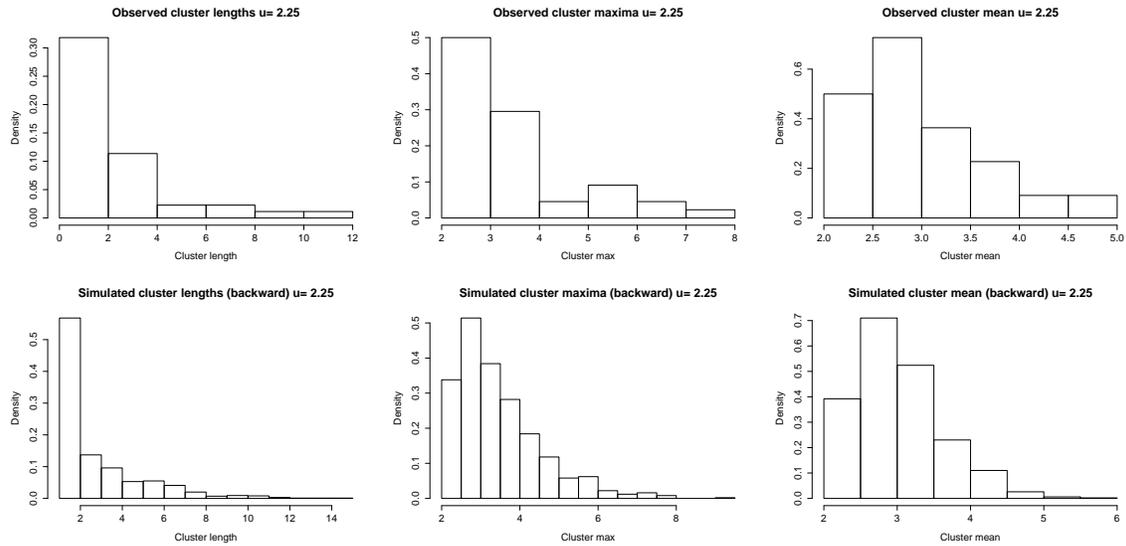
Figure 31: Observed and backwards simulated cluster characteristics with threshold $u = 2.25$
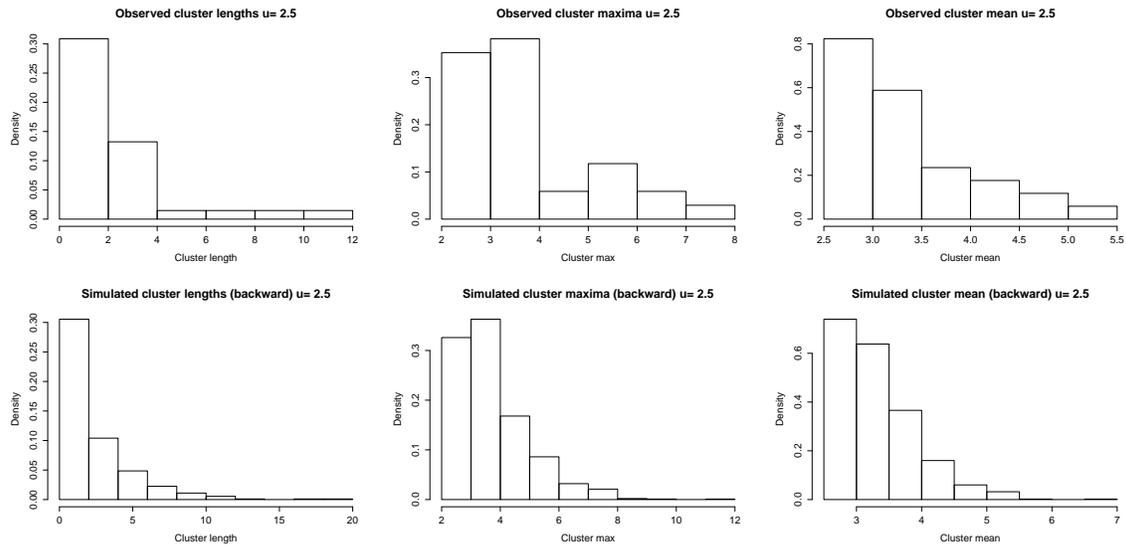
Threshold $u = 2.5$

Figure 32: Observed and backwards simulated cluster characteristics with threshold $u = 2.5$

The characteristics of the simulated data now appear much more consistent with those of the observed sample. The backwards simulation scheme did not have the same effect on the more symmetric Dalton Holme data, where the second order forwards scheme appeared to work better.

	Mean cluster length	Mean cluster maximum	Mean cluster mean
Observed	2.76 (0.44)	3.75 (0.23)	3.29 (0.12)
Simulated	2.82	3.76	3.30

Table 9: Observed and backwards simulated cluster characteristics with threshold $u = 2.5$

4.4.4   Discussion

For Chilgrove House, neither the first nor second order simulation schemes captured the characteristics of the clusters well. The asymmetry in this dataset (which may be due to covariates that are unaccounted for) perhaps means that the assumptions are too simplistic. In particular, there are hints of a mixture structure in the right-hand panel of Figure 17. It seems plausible that given a large value of X_t, X_{t+1} may come from a distribution that depends on covariates (e.g. rainfall), so that X_{t+1} is again large if there is low rainfall, and small if there is high rainfall. There may be a physical explanation for this observation in that normally a GW drought would not break immediately as the first rainfall would need to wet up the recharge pathways before GWLs would respond. However, under extreme rainfall other quick recharge pathways are thought to become active leading to rapid bypass of the unsaturated zone and under these circumstances then GWLs may respond more quickly to larger / more intense episodes of rainfall. This has not been fully investigated, but the left-hand panel of Figure 33 displays the one-step forwards residuals from Chilgrove House against rainfall, indicating the type of negative relationship that might be expected. On the other hand, since drought takes a long time to build up, the backwards direction has a simpler structure: if X_{t+1} is large (i.e. low groundwater), then it is likely that X_t is also large.
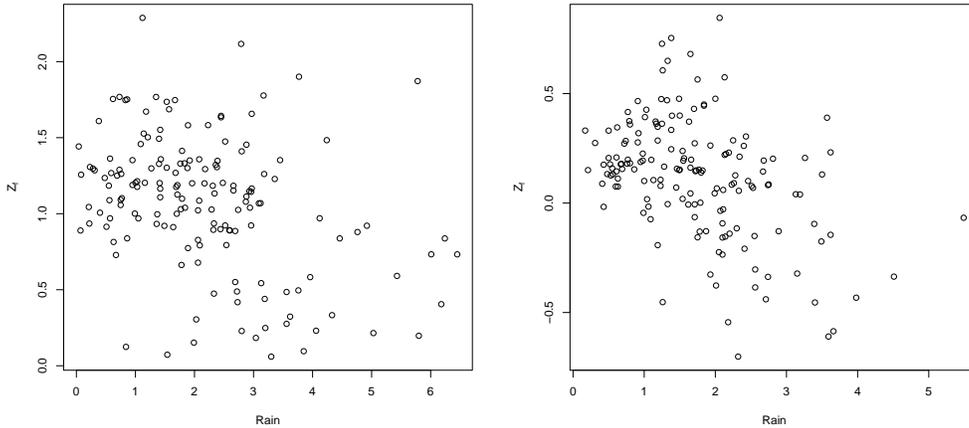
Figure 33: One-step forwards residuals against rainfall. Left: Chilgrove House, Right: Dalton Holme

The data at Dalton Holme appear more symmetric (though the right-hand panel of Figure 33 suggests some relationship between forwards residuals and rainfall). As a consequence forwards and backwards simulation schemes are more similar. A first order forward scheme overestimated cluster length and characteristics such as maximum and mean value. The proposed modification, which uses conditional sampling of residuals to take into account negative dependence between $(X_t, X_{t+2})|X_{t+1}$, appears to generally capture the observed characteristics, although with some

small overestimation remaining.

4.5 Case Study Summary

Through the course of this case study LU have demonstrated a number of extreme value theory methodologies that are applicable to the study of extreme GWL and other phenomena that are of interest to BGS such as volcanoes, space weather events and soil contamination. These methodologies can be used to predict the likelihood of the occurrence and severity of extreme events as a function of available covariates. Thus, in the case of GWL extremes it is possible to compare the current frequency of extremes with the frequency that might occur according to future climate scenarios. Other models can be used to simulate successive extreme GWLs and to determine properties of a multi-month extreme events. These models are more flexible and generally applicable that the methodologies currently used by groundwater scientists when considering groundwater extremes. The application of these models does require assumptions about the variable of interest but LU have demonstrated strategies that might be used to relax these assumptions when they are inappropriate. The methodologies applied are analogous to the methods currently used by BGS in that the GWLs are standardised and deviations from the seasonal norm can be identified in terms of thresholds on these standardized variables. However, these extreme value theory techniques acknowledge the potential for different behaviours in the tails of the distribution. BGS are now applying these methodologies to a wider set of GWL data and in collaboration with LU will produce a paper assessing the potential to use extreme value theory methodologies to characterise extreme groundwater events.

4.6 References

Adams B., Bloomfield J.P., Gallagher A.J., Jackson C.R., Rutter H.K., Williams A.T. 2010. An early warning system for groundwater flooding in the Chalk. Q. Journal Eng. Geol. and Hydrol., 43, 185-193

Bloomfield J.P., Marchant B.P., 2013. Analysis of groundwater drought using a variant of the Standardised Precipitation Index. Hydrology and Earth Systems Science, 17, 4769—4787.

Eltahir, E.A.B., Yeh, P. J-F., 1999. On the asymmetric response of aquifer water level to floods and droughts in Illinois, Water Resources Research, 35, 1199-1217.

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

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

Mackay, J.D., Jackson, C.R., Wang, L., 2014. A lumped conceptual model to simulate groundwater level time-series. Environmental Modelling and Software, 61, 229—245.

McKee T.B., Doesken N.J., Kleist J. 1993. The Relationship of Drought Frequency and Duration to Time Scales. Proceedings of the Eighth Conference on Applied Climatology. American Meteorological Society: Boston; 179-–184.

Monteith J.L., Unsworth M.H. 2008. Principles of Environmental Physics: Third Edition. Elsevier, London, UK.

Smith, R.L., Weissman, I., 1994. Estimating the extremal index. Journal of the Royal Statistical Society B, 56, 515–528.

Wilby, R.L., Noone, S., Murphy, C., Matthews, T.K.R., Harrigan, S., Broderick, C., 2016. An evaluation of persistent meteorological drought using a homogeneous Island of Ireland precipitation network. International Journal of Climatology, 36, 2854–2865.

Winter, H.C., Tawn, J.A., 2016. Modelling heatwaves in central France: a case study in extremal dependence. Journal of the Royal Statistical Society, C.

5 Appendix: MSc Project Proposals

5.1 The influence of soil texture on the potential for carbon sequestration

Numerous statistical challenges arise when mapping of soils, especially when assessing large areas and several properties that are driven by numerous controlling factors of various origins and scales. The objective of this study will focus on the development of a statistical framework to map together a set of soil properties in a 3 dimensions geographical space. Most of the studies based their maps on a model for one depth and for a single property. This may produce inconsistent predictions of the set of variable at a given depth. For example, clay is one of the driving factors of organic carbon. Both are major input parameters of climate or agronomic models. They should then be predicted together The study area will be the "Centre" region of France corresponding approximately to the middle Loire catchment) and has a surface area of about 39472 km2. Two datasets will be used. The first soil dataset A consists of 2487 soil profiles of the French soil database (IGCS - Inventory Management and Soil Conservation) coming from the historical soil database and from the campaign related to the realization of the regional 1:250k soil map. The second dataset B is extracted from the database coming from the French soil monitoring network (Réseau de Mesures de la Qualité des Sols: RMQS) (Jolivet et al., 2006) which is based on a systematic random grid (16 km x 16 km). The raw observations correspond to measurements realized on samples coming from soil horizons. The depths vary between sites. The modelling procedure will focus on a set of 4 soil properties: particle size (sand, silt, clay) and carbon. Additive-log-ratio (alr) transform will have to be applied to particle size data in order to keep the compositional constraint thorough the modelling procedure. We will use auxiliary environmental covariate such as digital elevation model or geological maps in the framework of the linear mixed model. Dataset A will be used as a calibration dataset and dataset B as a validation dataset. The predictions will be produced per depth increments that are the same at each prediction sites

5.2 Inter-relationships between rainfall, river level and ground water level

Groundwater level is a measure of the amount of water in storage within an aquifer. Understanding how groundwater levels fluctuate over time is important as abstraction of this water is often required, e.g. to supplement reservoir supplies in times of drought. Scientists at the British Geological Survey (BGS) have available high-frequency data on groundwater levels (every minute), river levels (every minute) and rainfall (every 15 minutes) for a site in Wallingford, Oxfordshire. The goal of this project will be to understand the key drivers behind fluctuations in the groundwater level at this site. Whilst it is clear that the groundwater level will be driven by both rainfall and river level, it is much less clear what the most informative lag time will be to relate these two processes to the groundwater level. Further this lag will probably differ between river level and rainfall. To investigate this, you will use exploratory data analysis to tease out the scale dependencies amongst

these three, highly related, series. You will then look at how this information can be incorporated into either a regression or multivariate model in order to fully quantify the relationships identified.

5.3   Extreme Space Weather

This project will be based on previous work by Alan Thompson and co-authors, who posed the question 'How extreme can geomagnetic storms become'? Geomagnetic storms have the potential to impact on the UK power grid, aviation, satellite operations, GPS and communications and therefore the UK government are interested in quantifying the 1-in-100 and 1-in-200 year extreme events for such storms. Geomagnetic time-series data is available for 28 European observatories. The data is very high frequency (1-minute) and so some declustering is required prior to the application of a peaks over threshold extreme value model. Existing work on this data has modelled the data across sites separately, and has also ignored any non-stationary features in the data. Your core aim will be to investigate any patterns and trends that may be observable within the extreme events, be these trends in space or time. Initially you will conduct a thorough exploratory data analysis, and then use the results of this to construct an appropriate temporal, or spatial, model.