Reconstruction of multi-decadal groundwater level time-series using a lumped conceptual model

C. R. Jackson[1], L. Wang[1], M. Pachocka[1], J. D. Mackay[1], J. P. Bloomfield[2]

[1] Environmental Science Centre, British Geological Survey, Keyworth, Nottingham, NG12 5GG, UK

[2] British Geological Survey, Wallingford, Oxfordshire, OX10 8BB, UK

Abstract

Multi-decadal groundwater level records, which provide information about long-term variability and trends, are relatively rare. Whilst a number of studies have sought to reconstruct river flow records, there have been few attempts to reconstruct groundwater level time-series over a number of decades. Using long rainfall and temperature records, we developed and applied a methodology to do this using a lumped conceptual model. We applied the model to six sites in the UK, in four different aquifers: Chalk, limestone, sandstone and Greensand. Acceptable models of observed monthly groundwater levels were generated at four of the sites, with maximum Nash – Sutcliffe Efficiency scores of between 0.84 and 0.93 over the calibration and evaluation periods, respectively. These four models were then used to reconstruct the monthly groundwater level time-series over approximately 60 years back to 1910. Uncertainty in the simulated levels associated with model parameters was assessed using the Generalized Likelihood Uncertainty Estimation method. Known historical droughts and wet period in the UK are clearly identifiable in the reconstructed levels, which were compared using the Standardized Groundwater Level Index. Such reconstructed records provide additional information with which to improve estimates of the frequency, severity and duration of groundwater level extremes and their spatial coherence, which for example is important for the assessment of the yield of boreholes during drought periods.

Introduction

Good quality, multi-decadal records of groundwater levels are important because they enable a better understanding of natural variability and trends in groundwater levels, including the impacts of recent climate changes (Chen et al., 2004; Hanson et al., 2006; Chen and Grasby, 2009; Little and Bloomfield, 2010; Holman et al., 2011; Stoll et al., 2011; Jackson et al., 2015). In addition, they support more accurate assessments of the frequency, persistence and severity of extreme groundwater levels, which inform water resource and risk planning (Rutulis, 1989; Butterworth et al., 1999; Peters et al., 2006; Bloom field and Marchant, 2013). However, such long-term groundwater level records are relatively rare. In most parts of the world short groundwater level records, of at best a few decades, are the norm, with hardly any records starting prior to the mid-20th century (Jousma and Roelofsen, 2004). Consequently, there is a need for robust and flexible methods to reconstruct groundwater levels to lengthen relatively short observational records.

There is a mature literature on the reconstruction of river flows from climate data, refer for example to Jones (1984), Jones and Lister (1998), Jones et al. (2006), Bartl et al. (2009) and Wen (2009), and references therein. In contrast, there have been few previous attempts to reconstruct groundwater levels (Schilling and Einfalt, 1982; Chelmicki et al., 2002; Ferguson and George, 2003; Conrads and Roehl, 2007; Najib et al., 2008; Perez-Valdivia and Sauchyn, 2011). Groundwater studies have either used a variety of statistical or neural network models that typically relate groundwater levels to

climate records or proxies for climate records. Statistical or process-independent methods such as neural network models have been favoured because, for many sites, the development of physically based, process models (such as distributed groundwater models) based on a full water balance requires more hydrogeological, hydrological and climate data than are often available to condition the models and also needs a level of conceptual understanding of the system that is not easily achievable.

Schilling and Einfalt (1982) reconstructed approximately 5 years of groundwater level data. They used a multivariate stochastic linear model of groundwater level as a function of internal persistence, external hydrological variables (precipitation and actual evaporation), artificial withdrawal and stochastic disturbances. Ferguson and George (2003) used a stepwise multiple linear regression as a function of temperature, precipitation and tree ring width to reconstruct mean annual groundwater levels. First, a principal component analysis was performed to explore if the hydrometric networks of interest contained any sub-groups related to location or hydrogeological setting. This identified two sub-groups. The groundwater levels of these groups were then correlated against seasonal and annual instrumental climate (temperature and precipitation) and tree ring data for lags of zero to five years. Using a 32-year calibration period their model was able to explain ~72% of the variance. They then used the calibrated model to reconstruct annual groundwater levels between 1907 and 1965. Chelmicki et al. (2002) used artificial neural networks (ANNs) to reconstruct monthly groundwater levels between 1901 and 1960 using air temperature and precipitation data. Conrads and Roehl (2007) also used ANN models to reconstruct hourly groundwater levels. Groundwater level records from observation boreholes close to sites where groundwater levels were being reconstructed were used to train the ANNs. The ANNs were effective in modelling groundwater levels over the calibration period and were used to reconstruct groundwater levels up to periods of eleven years.

Najib et al. (2008) reconstructed daily groundwater levels over a 40-year period to assess the frequency of extreme groundwater levels and associated flooding caused by groundwater discharge from a fractured carbonate aquifer in south-east France. They developed a model composed of two modules based on simple time- stepping functions. The first is a soil water accounting algorithm that relates potential recharge to rainfall, potential evapotranspiration (PET) and soil storage. The resulting recharge time-series was used in the second module, which calculates the groundwater level based on linear storage, and non-linear discharge, groundwater head-dependent functions. The model has five non-physically based parameters, and for the case study borehole was calibrated against observed levels for a one- year period, September 2002 to October 2003, achieving a Nash – Sutcliffe Efficiency (NSE; Nash and Sutcliffe, 1970) of 0.74. It was then applied to reconstruct post- 1960 levels, which are used for flood frequency analysis. The model fi t to the observed groundwater levels is reasonable over the one-year calibration period. However, there was no assessment of its performance over a separate evaluation period, even though it was used to simulate extremes in a fractured, karstic aquifer, and no assessment of model uncertainty was undertaken.

Perez-Valdivia and Sauchyn (2011) used tree ring chronologies to reconstruct mean annual groundwater levels in two observation wells in Alberta, Canada. They developed step-wise multiple linear regression models based on predictors formed from the five most correlated chronologies and achieved reasonable fits to the 36- and 42-year observed records; models with adjusted R^2 values of 0.71 and 0.47 were achieved. Whilst the two records were modelled over the relatively long periods of 354 and 105 years, this approach only allowed annual levels to be reconstructed. None of these previous models were evaluated against long observational records, so their performance outside the calibration period was untested. In addition, measures of confidence associated with the estimates of reconstructed groundwater levels are generally not available. We addressed these limitations in this study and present a methodology for reconstructing monthly multi-decadal groundwater levels, in our

case back to 1910, from shorter groundwater level observational records and longer climate records. To do this we used a conceptualized lumped aquifer model driven by monthly rainfall and PET time-series. We applied the model to six sites in England across five of the major aquifers types within the UK: Permo-Triassic Sandstone, Magnesian and Jurassic Lime- stone, Lower Greensand and Chalk (Allen et al., 1997). Each model was calibrated and evaluated against observed groundwater levels, prior to its use for reconstruction. Calibration was performed through a Monte Carlo simulation process, and multiple acceptable or behavioural models were used to place uncertainty bounds on the reconstructed groundwater level time- series using the Generalized Likelihood Uncertainty Estimation (GLUE) methodology (Beven and Binley, 1992). We tested the methodology using one observed groundwater level record that extends back to 1836, assuming that we did not have pre-1970 levels. The results of the groundwater level reconstructions are discussed in the context of known drought and flood periods before the period of groundwater level observations, prior to conclusions being made regarding the further application of the approach.

Model description

To simulate and reconstruct groundwater levels we developed lumped groundwater models using the AquiMod code, the development and application of which are presented in detail by Mackay et al. (2014). AquiMod models groundwater level time-series at observation boreholes by linking three simple algorithms that simulate soil drainage, the transfer of water through the unsaturated zone and groundwater flow (Figure 1). It takes rainfall and PET time-series as input, and in this study we used a monthly time-step, although this can be adjusted. The soil module partitions rainfall between evapotranspiration, runoff and soil drainage, based on a four-parameter soil moisture accounting procedure. A Weibull distribution transfer function is then used to attenuate soil drainage through the unsaturated zone, over a number of months, *n*, before it reaches the saturated zone as groundwater recharge. This approach is similar to that applied by Calver (1997), in which a proportion of the soil drainage in each month is applied to the water table over the current month and a number of subsequent months. The Weibull function can represent exponentially increasing, exponentially decreasing and positively and negatively skewed distributions. It is used because it allows the exploration of different distributions, whilst being smooth, which is considered to be more physically justifiable than randomly selected monthly weights. Furthermore, this method requires only three, rather than the n+1 model parameters of Calver (1997). Outflows from the saturated zone are calculated using Darcian flux equations based on the aquifer hydraulic conductivity and the difference between the water table elevation and the elevations of the saturated zone outlets (Figure 1).

AquiMod was selected for this study because it has been designed specifically for simulating groundwater levels at observation boreholes. It includes in-built Monte Carlo parameter sampling, it is fast to run and it has been shown by Mackay et al. (2014) to be able to simulate irregular groundwater hydrographs in a range of hydrogeological settings, for example in unconfined aquifers where hydraulic properties vary with depth.

Model application

Reconstruction sites. We applied AquiMod to reconstruct groundwater levels at six sites across England (Figure 2; Table I). These were selected to cover a range of the major aquifer types found in the UK. Individual boreholes were chosen based on the following criteria:

1. The groundwater levels are not significantly controlled by surface water levels and are indicative of bulk aquifer storage.
2. Groundwater abstraction in the catchment is small, and associated impacts on levels are likely to be insignificant.
3. At least 20 years of monthly groundwater level measurements were available.

The first of these criteria was assessed by examining the hydrogeological setting of the borehole and the pattern of groundwater level fluctuation. It is generally clear whether variability in observed groundwater levels is controlled by the stage of a nearby river, or whether levels predominantly vary in response to recharge, around a mean controlled by a distal surface discharge point. For example, the strong control of a river on a groundwater level record is often apparent when the annual maximum or minimum series varies little. This is not the case at the sites considered in this study. In the UK all groundwater abstractions pumping at an average rate greater than 20 m^3day^{-1} require a licence. The second criterion was assessed using groundwater abstraction licence information provided by the regulator, the Environment Agency, as part of an earlier study that assessed the impact of future climate on UK groundwater levels (Prudhomme et al., 2013).

The six sites are as follows:

Bussels No. 7A: A 91 m-deep borehole drilled through 48 m of river terrace deposits into the Permian Dawlish Sandstone Formation in south-west England. The construction of the borehole is unknown but is likely to be cased out through the river terrace deposits. Monthly observed groundwater levels are available from November 1971 and have fluctuated between 22.9 and 25.3 m above sea level (a.s.l.) or 3.4 and 1.0 m below ground level (b.g.l.). The range of fluctuation in levels is associated with the specific yield of the overlying sands and gravels for which typical values will be in the range 5 – 20%. Estimates of transmissivity derived from pumping tests in this aquifer (and for other sites), reported by Allen et al. (1997), are presented in Table I as are typical values of specific yield.

Chilgrove House: A 62 m-deep borehole drilled into the unconfined Cretaceous Chalk aquifer in the catchment of the River Lavant in south-east England. Approximately monthly measurements of groundwater level have been taken since 1835 (we use this site as an evaluation of our methodology). Annual fluctuations are usually around 20 to 30 m, but monthly rises in level have exceeded 34 m. Transmissivity and storage development in the Chalk is associated with the widening of fractures by dissolution, which is more significant in the zone of water table fluctuation; typically there is a non-linear decrease in transmissivity and storage coefficient with depth (Williams et al., 2006).

Lower Barn Cottage: The depth and construction of this borehole are unknown. However, it has been used to monitor groundwater levels in the unconfined Cretaceous Lower Greensand Group, an important aquifer in south-east England, since 1975. The Group comprises a complex series of variably cemented clays and sands, which are approximately 9 m thick in this area. The heterogeneity of the aquifer produces a relatively irregular hydrograph, the annual mean of which is only 2.6 m b.g.l.

New Red Lion: This site is located on the Jurassic Limestone in central England, a relatively fast responding fracture-dominated aquifer. The groundwater level time-series at this site represents the amalgamation of two records: that of the 50 m-deep Old Red Lion borehole, which was used to monitor groundwater levels from 1964 and 1981, and the New Red Lion borehole, of similar depth, 94 m away, from which the 1981 – present record is derived. The borehole is located in a valley which is incised through the Oxford Clay Formation into the underlying limestone sequence. The groundwater level at the borehole varies from unconfined conditions, at minimum water levels, to

con fi ned conditions at maxi- mum water levels. The degree of local aquifer confinement varies temporally and spatially. The hydrograph has an annual sinusoidal response with a seasonal variation generally of about 8 to 12 m.

Skirwith: This 89 m-deep borehole has been used to monitor groundwater levels in the Triassic St Bees Sandstone in Cumbria, north-west England, since November 1978. At this location the aquifer is con fi ned. However, superficial glacial till cover is areally discontinuous and groundwater level fluctuations are driven by direct rainfall recharge over uncovered, unconfined sandstone. The groundwater level varies from a maximum level (131.6 m a.s.l.) lying within overlying superficial deposits to a minimum level (129.4 m a.s.l.) lying in the sandstone. The hydrograph exhibits an annual sinusoidal response but with many smaller peaks.

Swan House: A 60 m-deep borehole penetrating the Permian Magnesian Limestone with a continuous groundwater level monthly record dating back to September 1973. The limestone is overlain by 26 m of glacial till, which confines the aquifer. Observed groundwater levels have fluctuated between approximately 78 – 90 m a.s.l. (17 – 5 m b.g.l.). Pumping test derived storage coefficients range over several orders of magnitude from 3.4×10^{-6} to 2.4×10^{-2} with an interquartile range of 1.3×10^{-4} to 8.0×10^{-4} , although specific yield values are typically around 5% (Table I).

Driving climate data. To calibrate and evaluate the AquiMod models against the observed groundwater level time-series, and to use them to reconstruct groundwater levels, rainfall and PET time-series were required for each catchment. In this study we used monthly climate time- series to simulate monthly groundwater levels. Monthly climate data were derived from the UKCP09 data set (Jenkins et al., 2008). This has been generated by the UK Met Office (Perry and Hollis, 2005) and provides time- series for a range of climatic variables on a 5 km grid across the UK, based on the archive of UK weather observations. Gridded values have been generated using regression and interpolation from the irregular station network, considering co-variates such as latitude and longitude, altitude and terrain shape, coastal influence and urban land use. This moderates the effect of changes in the station network on homogeneity, but the impacts of station commissioning and closure cannot be removed entirely (Perry and Hollis, 2005). For temperature variables the number of stations rose from about 270 in 1914 to 600 in the mid-1990s, before falling to 450 in 2006. For precipitation the number of stations, for which monthly data are available, increases from approximately 200 in 1910 to 500 in 1914. In the early 1960s the number of rain gauges in the UK monitoring network jumped from 800 to 4600. The number peaked at 5700 in the early 1970s, before declining to 3000 in 2010 (Legg, 2011).

The UKCP09 data provide monthly rainfall totals between 1910 and 2012, and we used the associated time- series for the 5-km grid cell in which the borehole is located. However, PET series are not provided. The physically based Penman–Monteith equation (Monteith, 1965) is recommended by the United Nations Food and Agriculture Organization for deriving PET (Pereira et al., 1999). This is used in the UK Meteorological Office Rainfall and Evapotranspiration Calculation System (MORECS) (Field, 1983), which outputs a 40 × 40 km gridded monthly average PET dataset for the United Kingdom, between 1961 and present, based on synoptic station data and a modified version of the Penman–Monteith equation (Monteith and Unsworth, 2008). However, records of the weather variables required for the Penman–Monteith equation are not systematically available prior to 1961. Consequently, we used the Blaney–Criddle method (Blaney and Criddle, 1950), which only requires a temperature input, to convert the UKCP09 monthly mean temperature data into PET. An assessment of the uncertainty in simulated groundwater levels because of errors in derived PET was beyond the scope of this study, but Kay et al. (2013) review a number of studies that compare the calculation of PET using equations of varying complexity and their use in hydrological modelling. They state that because PET is much less spatially and temporally variable than rainfall, relatively simple data are often

sufficient to close a water balance in hydrological modelling. For example, they cite the study of Oudin et al. (2005), which evaluated 27 PET formulations applied to four rainfall-runoff models and more than 300 catchments in Australia, the USA and France. This study found that PET based on temperature or radiation often resulted in more accurate stream flow simulations than those using more complex formulae. However, given the very common use of MORECS PET in UK-based hydrological applications we bias corrected the Blaney–Criddle PET time-series against MORECS data using the equidistant quantile matching approach of Li et al. (2010a). This approach has an advantage over the traditional quantile matching approach in that it preserves any non-stationarity in the data that might have occurred over time. R-squared values calculated by comparing the resulting Blaney–Criddle PET and MORECS data over the period 1961–2012 are greater than 0.95 at all six sites.

We recognize that the use of monthly climate data to drive a soil-moisture accounting model, which runs on a monthly time-step, is known to underestimate potential recharge (Rushton, 2003). However, we have been restricted to this approach because appropriate daily time-series that could be used for reconstruction have not been available. Given that our aim was not to accurately quantify recharge but to simulate monthly groundwater levels, this was considered to be acceptable.

Calibration and evaluation. Rather than calibrating AquiMod through an optimization procedure that searches for a best model, we applied a Monte Carlo process to identify sets of acceptable or behavioural models. Indeed, the concept of a best model is erroneous, considering that parameter values are inherently uncertain, and given considerations of equifinality as discussed by Beven (2006). Consequently, multiple sets of model parameters were sampled from user-de fi ned ranges to generate an ensemble of simulations at each site.

Mackay et al. (2014) provide a detailed description of AquiMod, its application and parameterization, and we refer the reader there for a comprehensive description of its use. They also apply AquiMod to three of the sites considered in this study: Chilgrove House, Lower Barn Cottage and Skirwith. Consequently, we provide here a more concise description of the approach to AquiMod ' s parameterization. Values for 16 AquiMod parameters must be de fi ned (Table II) before a simulation can be run. All of these could be treated as calibration parameters; however, we fixed the values of seven of them using available catchment information. The catchment length, Δ x, was specified as the length between the observation borehole and a single discharge point on a neighbouring, down-groundwater flow gradient river based on an assessment of catchment geometry and hydrogeology. The catchment length and hydraulic conductivity parameters, K_i , are used in the calculation of discharges, Q_i , through the outlets of the saturated zone model component (Figure 1) based on Darcy's law and an equation of the form:

$$Q_i = K_i B_i \frac{\Delta h_i}{0.5\Delta x}$$

where B_i is an appropriately calculated saturated thick- ness, which depends on the current groundwater level and the elevation of the outlet being considered (Mackay et al., 2014). The catchment length parameter could be discarded by combining it with each of the hydraulic conductivity parameters. However, it is retained in AquiMod to enable users to think more easily about the physical system and to maintain the use of hydraulic conductivity, a basic hydrogeological parameter. Our approach to calibration is generally to fi x the catchment length parameter to a reasonable value based on an examination of the groundwater catchment size associated with the borehole and the distance between the borehole and catchment discharge points. Mackay et al., 2014 discuss the interaction and calibration of the catchment length and hydraulic conductivity parameters further. Marsh and Hannaford (2008) provide values of catchment base flow index, and Boorman et

al. (1995) present field capacity and wilting point values for UK soils. By analysing cross correlations between rainfall and ground- water levels, the n parameter of the unsaturated zone model component was set to the period over which there is a significant correlation at a 95% confidence level. The bottom outlet was set to the known bottom elevation of the aquifer based on geological and hydrogeological records held by the British Geological Survey (British Geological Survey, 2015). As Mackay et al. (2014) found, preliminary model runs showed that the upper two outlet elevation parameters significantly interacted with the hydraulic conductivity parameters. As such, a preliminary set of calibration runs were undertaken to determine elevation values for the middle outlet of each model that produced behavioural simulations to which they were subsequently set.

The remaining nine parameters were calibrated. The elevation of the upper outlet was varied randomly within the zone of water table fluctuation. Values for the other eight parameters were also randomly sampled from uniform distributions with upper and lower bounds specified based on knowledge of the hydrogeology of the aquifers. Allen et al. (1997) provide estimates of hydraulic property values for UK aquifers, and this was used to constrain parameter ranges. Values for the parameters that were fixed a priori are given in Table II, as are the ranges of the parameters that were found to be behavioural over the calibration periods. To identify acceptable models we assessed performance over both a calibration and evaluation period. Except for the Chilgrove House site, the calibration and evaluation periods were de fi ned by splitting the observed ground- water level record approximately in half (Table III). The half of the record with the greatest range in groundwater levels was then selected as the calibration period. For Chilgrove House, the period 1970–2012 was split to de fi ne the calibration and evaluation periods. This site has an observed record that starts in 1836, and therefore, it was not actually necessary to reconstruct its post-1910 groundwater levels. However, we included it in this study as it provides a further evaluation of the approach.

To assess uncertainty in the simulated groundwater levels over the calibration, evaluation, and subsequently the reconstruction periods we used the GLUE methodology of Beven and Freer (2001). This was first applied in a hydrological model application by Beven and Binley (1992) but has been applied to a wide variety of hydrological and environmental modelling applications since then (e.g. Freer et al., 1996; Piñol et al., 2005; Viola et al., 2009; Shen et al., 2012; Breinholt et al., 2013). In the GLUE approach the uncertainty in predictions is estimated using a set of behavioural models, which are weighted according to a likelihood measure describing how well they performed during calibration. GLUE is related to formal Bayesian analysis methods but implements an informal, or subjective, likelihood measure, rather than a formal model of the errors. To apply the GLUE methodology three assumptions must be stated explicitly: (i) what are the prior distributions from which the uncertain model parameters are sampled, (ii) which informal likelihood measure will be used; and (iii) what is the threshold value of this likelihood measure that differentiates behavioural and non-behavioural models. Given little information on prior distributions of calibration parameters, uniform distributions were selected in all cases in this study. Similar to many other applications of GLUE to hydrological models we adopted the NSE (Nash and Sutcliffe, 1970), as the likelihood measure. We selected a NSE value of 0.5 to represent the cut-off between non-behavioural simulations, which are discarded, and behavioural simulations. Given model run-time constraints, 10^6 simulations were performed within each Monte Carlo run at a site. A significant number of behavioural models (Table III) were obtained at the four non-limestone sites where between 3.3 and 9.3% of the simulations were behavioural. However, at the two limestone sites, New Red Lion and Swan House, only 0.12 and 0.18% of the simulations produced NSE values greater than 0.5, respectively. Consequently, the Monte Carlo run was repeated for the limestone aquifers, and the number of simulations increased to 10^7 . The highest NSE obtained over the calibration period ranged between 0.77 and 0.93 across the sites, with the sandstone and Lower Greensand models performing best

(Table III). The simulated time-series over the calibration period are plotted in Figure 3. In addition to the observed data (solid lines), the hydrographs simulated by the model with the highest NSE (dashed lines), and the envelope of all behavioural models, are plotted.

At all sites the ensemble of behavioural models, as represented by the GLUE 5 and 95% likelihood-weighted prediction limits, performs well in bracketing all of the observations during the calibration period, except at New Red Lion, where the levels around the annual minima in 1965 and 1976 are not captured. In general, the models reproduce the inter-annual and multi-annual variability wells, particularly for example, at Lower Barn Cottage, which has a more irregular hydrograph. The major drought of 1976 is simulated particularly well by the best model of the Bussels No. 7A hydrograph. As an additional quantitative measure of the performance of the models during both the calibration and evaluation periods, we also calculated the containment ratio (CR) (Xiong and O ' Connor, 2008). This index describes the proportion of observed values that are enclosed by chosen lower and upper GLUE likelihood-weighted prediction limits. Examples of its use in hydrological modelling studies include those of Xiong and O ' Connor (2008), Li et al. (2010b), Franz and Hogue (2011) and Breinholt et al. (2013). We calculated the CR for each site using the 5 and 95% GLUE prediction limits, and for various NSE likelihood threshold values of 0.4 and above. These are listed in Table IV and plotted in Figure 4, which shows the difference in performance and the level of uncertainty between the non- limestone and limestone models over the calibration periods; CR values are less than 68% with a NSE likelihood threshold of 0.5 at the New Red Lion and Swan House sites, whereas CR values are all significantly higher for the other sites for this likelihood threshold. The purpose of calculating the CRs for different likelihood thresholds was to examine the impact that the selected threshold had on the uncertainty limits. The CR curves level off as the NSE threshold applied in GLUE reduces. Based on this plot the widely applied threshold of 0.5 (e.g. Mo et al., 2006; Mittman et al., 2012; Shen et al., 2012), which was selected a priori, was considered reasonable.

All of the behavioural models over the calibration period were used to simulate groundwater levels during the evaluation period, and the NSE recalculated. The maximum NSE reduced slightly between the calibration and evaluation periods (Table III) at all sites except Chilgrove House, where it increased from 0.87 to 0.93. The number of behavioural simulations reduced between calibration and evaluation runs at all sites by between 43 and 75%. The GLUE prediction limits were calculated over the evaluation period (Figure 3) by using (i) all of the behavioural models over the calibration period and (ii) only those that achieved a NSE of 0.5 or above over the evaluation period. In addition to the observed data over the evaluation period, the median of the distribution of the likelihood-weighted time-series (dashed lines) is plotted in Figure 3. As for the calibration periods, the AquiMod models reproduce the inter-annual and multi-annual variabilities of the observed records over the evaluation periods well. The most significant low and high groundwater level events in the UK, in the period 1970–2011, occurred during the drought of 1976 and the extremely wet winter of 2000/2001. It is notable that these events are simulated well by the models when they occur in the evaluation period. However, it is clear that the periods of lower than average groundwater level in the New Red Lion record during the 1990s are not simulated adequately. During the calibration process a number of tests were performed to attempt to improve the performance of the models of New Red Lion and Swan House. This included widening the sampled parameter ranges but this did not result in higher NSE values over the calibration periods. As Mackay et al. (2014) found AquiMod has not been able to reproduce groundwater levels as well at limestone sites. We attribute this to the simplified representation of the real vertical hydraulic conductivity and storage structure in these fracture dominated aquifers, which for example results in a flashy response to recharge when groundwater levels are high. We have since found that increasing the temporal resolution of the model from a

monthly to a daily time-step significantly improves the simulation of extreme groundwater levels in limestone aquifers, but given access to monthly driving data only were not able to undertake this.

CR values for the evaluation periods are given in Table IV. For the four non-limestone sites the CRs are all above 79%. For New Red Lion and Swan House, the CRs are 64.0 and 66.8%, respectively. The change in the CR value between the calibration and evaluation periods depends on the site. At Chilgrove House, New Red Lion and Swan House they differ by only as much as 4.5%. At Bussels No. 7A, Lower Barn Cottage and Skirwith they decrease by between 8.5 and 10.8%. In addition to calculating the proportion of all the observations contained by the 5 and 95% prediction limits over the evaluation periods, we also calculated CRs for the values in the lower and in the upper half of the observed distribution. This provides a measure of the relative performance of the models in simulating low and high levels. From these CR values, also listed in Table IV, it is apparent that low levels are captured better at three of the sites (Bussels No. 7A, Chilgrove House and Swan House) and high levels captured better at two sites (New Red Lion and Skirwith). Levels in the lower half of the distribution are simulated particularly well at Chilgrove House, for which the associated CR is 100%.

Reconstruction of groundwater levels. Based on the CRs and visual inspection of the ensemble of behavioural models over the evaluation period it was considered acceptable to use AquiMod to reconstruct the groundwater level records back to 1910 at all of the sites except New Red Lion and Swan House. Therefore, the models that achieved a NSE of 0.5 or above over the evaluation period were used to do this at the other four sites. All models were initialized by setting the groundwater level at the start of January 1910 to mean January levels. The end of the simulation period was the end of March 2012. The resulting simulated time-series are shown in Figure 5, in which the median of the ensemble of likelihood- weighted simulated levels (dashed lines) and the envelope of the 5 and 95% prediction limits (grey bands) are plotted, in addition to the observed time-series (solid lines).

Discussion

The reconstructed groundwater levels enable a comparison of extreme events in the observed and reconstructed record. Additional information about groundwater level minima is particularly important, for example for the assessment of borehole yields during drought, which in the UK must be estimated by private water companies to comply with Government legislation (Misstear and Beeson, 2000). During the winters of 2000 – 2001, 2012 – 2013 and 2013 – 2014 many permeable catchments across south-east England and northern France experienced severe groundwater flooding in areas where it had not previously been observed (Habets et al., 2010; Hughes et al., 2011). Placing these maxima within the context of the longer reconstructed record would also enhance groundwater flood risk assessments.

The major droughts in England and Wales since 1800 have been identified by Cole and Marsh (2006) and Marsh et al. (2007). These are summarized in Table V, in addition to the other less severe, but still significant, post- 1910 droughts that are described elsewhere (Royal Meteorological Society, 1948; Phillips and McGregor, 1998; Fowler and Kilsby, 2002; Lloyd-Hughes et al., 2009). The two most severe droughts between 1910 and 1970 are those of 1921–1922 and 1933–1934. The 1976 drought is taken as the benchmark event across much of England and Wales (Marsh et al., 2007), when flows in the majority of British rivers fell to their lowest recorded levels, and groundwater resources were severely impacted. The driest year on record over England was 1921 for which the annual rainfall total was approximately 570 mm (Met Office, 2014). The years 1933, 1964, 1973 and 1996 are the other four driest years within the rainfall record starting in 1910.

These historical droughts appear in the reconstructed groundwater level records but their signals differ between the sites (Figure 5). The 1933–1934 drought is identifiable in all four reconstructed time-series but is more prominent at the three southern sites. Similarly, the 1921–1922 drought is more clearly distinguishable in the time-series of Bussels No. 7A and Lower Barn Cottage, and of Chilgrove House where the model reproduces the groundwater level observations very well. This is consistent with Marsh et al.'s (2007) description that these two droughts were more severe across southern Britain. The droughts of the early and late 1940s, and 1962–1964, are also distinguishable. Groundwater levels are simulated as being extremely low in 1949 at all sites except Skirwith. Lower than average levels are simulated to persist during the early 1960s at all sites, but are only extreme compared with other episodes, in the Bussels No. 7A and Skirwith time-series. Because of a trend of decreasing levels from the start of the 1970s, groundwater levels at Lower Barn Cottage are lower in 1973 than during the 1976 drought, which are simulated to have recovered following relatively high rainfall in 1974. The droughts of the 1970s are simulated to be more persistent in the Skirwith time-series because of the associated high storage of the sandstone aquifer.

The historical record of individual flood events within the UK is good (Black and Law, 2004). However, there is not a systematic review of persistent wet periods in the UK in the hydrological literature, similar to those undertaken for droughts. The wettest year over England during the period 1910–2013 was 2012 for which the annual rainfall total was approximately 1125 mm (Met Office, 2014). The years 1912, 1960, 2000 and 2002 were the other four wettest years over this 114-year period.

Notable groundwater level maxima in the reconstructed records occur during 1947 and 1960. Records of severe flooding following rapid thawing of heavy snowfall in early 1947, and subsequent record rainfall in the March of that year (Risk Management Solutions, 2007), are consistent with the reconstructed groundwater levels. The third highest recorded rainfall total for England fell in 1960, and the simulated extreme groundwater level maxima in this year are consistent with this.

To allow direct comparison of the reconstructed time- series across the sites the monthly Standardized Ground- water level Index (SGI) (Bloomfield and Marchant, 2013) was calculated for the simulated 1910–2012 groundwater level time-series at each site. The SGI is a variant of the Standardized Precipitation Index (McKee et al., 1993; Edwards and McKee, 1997). It is estimated using a non-parametric normal scores transform of groundwater level data for each month. The monthly scores are then combined to form a continuous index. The SGI time- series are plotted in Figure 6 as a heat map, from which it is relatively straightforward to identify the spatial coherence of events, which events are more persistent, and the degree of autocorrelation in each series. The impacts of the 1933–1934, 1973, 1976, 1992 and 1995–1997 droughts (in red) are clearly distinguishable across all sites as are the notably wet periods (in blue) of 1947 and 2000–2001. The longer memory in the Skirwith and Lower Barn Cottage observation borehole time-series, which monitor levels in high storage sandstone and Lower Greensand aquifers, respectively, is apparent, for example compared with the Chalk hydrograph of Chilgrove House.

In addition to the SGI series for the reconstructed Chilgrove House record, the SGI series based on the observations is also shown. This again shows the good match between the reconstructed and observed levels at this site, although there are some small differences. For example, the 1933/1934 drought is more severe, and the wet period of 2000/2001 not as intense, in the reconstructed record.

The heat map highlights a feature of the modelled Skirwith hydrograph, which exhibits a period of higher levels between 1910 and the early 1930s. The question arises whether these modelled levels are accurate or whether they are an artefact of the modelling process, for example because of the availability of less, or poorer quality, climate data over this period. To explore this issue, without

undertaking a full assessment of the contribution of rainfall uncertainty to simulated groundwater level uncertainty, which is beyond the scope of this paper, we additionally plot the groundwater level time-series as level-duration curves (LDCs) in Figure 7. LDCs are plotted for the best model for each site for the calibration/evaluation and reconstruction periods. Additionally for the Chilgrove House site a LDC is plotted for the observed groundwater levels over the reconstruction period. Comparison of the simulated and observed LDCs for the calibration/evaluation period again indicates the generally good fi t between the modelled and observed data. However, levels lower than approximately that of the 90th percentile of the distribution at Lower Barn Cottage are not simulated as accurately. The LDCs for the simulated calibration/evaluation and reconstruction periods are similar at Bussels No. 7A and Lower Barn Cottage. However, there are larger differences between these LDCs at Chilgrove House and Skirwith. At Skirwith the LDC shows that the range of groundwater levels is higher over the reconstruction period as are groundwater levels in general. Similarly, the LDC for Chilgrove House shows that simulated groundwater levels are higher over the reconstruction period. Comparison to the LDC for observed groundwater levels over the reconstruction period indicates that model error is more significant over the reconstruction period at this site.

To assess if the reason for the difference in the LDCs at Chilgrove House and Skirwith is potentially because of differences in climate data, we plot kernel density plots (KDPs) for monthly rainfall and PET over the calibration/evaluation and reconstruction periods for the four sites in Figure 8. The shape of the two KDPs for PET are reasonably similar at each site. Mean PET is higher over the later calibration/evaluation period at all sites by between 2.3 and 3.4%, reflecting slightly warmer temperatures over the last 40 years. The difference in mean rainfall between the reconstruction and calibration/evaluation period is in the range 0.2 to 2.4% for all sites expect Chilgrove House, where the UKCP09 rainfall is 7.4% higher over the reconstruction period than over the calibration period. To test whether this higher rainfall could account for the simulated Chilgrove House levels being higher over the reconstruction period, we undertook a further simulation of the 1910–1969 period using an adjusted rainfall sequence. We bias-corrected the 1910–1969 rainfall using quantile mapping against the 1970–2012 series to generate a rainfall series with the same mean and distribution as that of the latter period. This produced simulated levels that are lower than those observed at all percentile points of the distribution (Table VI); the median reconstructed value decreased from 49.5 to 45.2 m a.s.l., compared with 46.8 m a.s.l. for the observed record.

To further examine the difference between the simulated and observed levels over the reconstruction period at Chilgrove House we plot this error as an average annual time-series in Figure 9a. The model error is smaller after 1961, when the number of observations from the rain gauge network within a 20 km radius of the borehole increased markedly (Figure 9c); 20 km is larger than the scale of the borehole catchments (Figure 2). It is clear from the KDP of rainfall and the change in the number of rainfall measurements over time that errors in the reconstructed levels at this site are in large part because of more limited rainfall data over the reconstruction period. Similar time-series of the number of rainfall measurements within a 20 km radius of the other three reconstructed sites are also plotted in Figure 9. It is likely that the paucity of rainfall measurement prior to 1961 at these other sites also introduces error into model, although this is not as easy to distinguish from a visual inspection of the level duration curves (Figure 7). Further work to assess how uncertainty in the rainfall data contributes to uncertainty in the reconstructed levels is required, particularly at the Skirwith site where the climate data distributions before and after the start of the observed groundwater level record are similar. This should include a consideration of sensitivities to both spatial and temporal (e.g. Price et al., 2014; Sapriza-Azuri et al., 2015) variability in driving climate data, but this was beyond the scope of this work.

Conclusions

Whilst there is increasing recognition of the bene fi t of long-term historical groundwater level observation to assess, for example future climate change within the context of past trends and variability, or the resilience of water resource systems to extreme events, few studies have proposed approaches to reconstruct historical groundwater levels. We have developed a methodology to do this, based on the use of a lumped conceptual groundwater model, AquiMod, driven by available gridded monthly rainfall and PET data, and applied it to six sites in the UK located on different aquifers. Because of its simplicity AquiMod runs quickly and has allowed the use of a Monte Carlo approach to calibrate and evaluate multiple models to historical observations and to assess model parameter uncertainty by using the GLUE method. Acceptable models of historical monthly ground- water levels were generated at four of the sites, with maximum NSE scores of between 0.84 and 0.93 over the calibration and evaluation periods. These models were then used to reconstruct levels back to 1910. Models of the two limestone sites, which have more irregular time- series because of the fractured nature of these aquifers, obtained maximum NSE scores of 0.70 and 0.77 over the calibration period. However, given the number of observations contained within the 5 and 95% GLUE uncertainty bounds (less than 67%) these models were not considered acceptable to use for reconstruction. Known historical droughts and wet periods in the UK are clearly identifiable in the reconstructed levels. At two of the sites (Chilgrove House and Skirwith) levels were simulated to be lowest in 1973 and 1976, which for Chilgrove House agrees with the observational data. At Bussels No. 7A the first and second ranked drought years are 1921 and 1934, the order of which is reversed at Lower Barn Cottage. Notable groundwater level maxima in the reconstructed record occur during 1947 and 1960, which correspond with knowledge about flood events during these years. Groundwater levels for Chilgrove House were reconstructed between 1910 and 1970 even though the observed record covers this period. This provided a further test of this model, which exhibited larger errors prior to 1961. This has been related to errors in the gridded rainfall dataset that was used, arising from the significantly lower number of rainfall measurements made within a 20km radius of this site prior to 1961. Further work is required to quantify the uncertainty of rainfall and PET gridded time-series data and to explore how this uncertainty propagates through to modelled groundwater levels. However, this was beyond the scope of this paper and will be the subject of future research.

Acknowledgements

We thank Tim Legg at the UK Met Office for information about historical changes to the density of the UK rain gauge network. This work has been funded by the British Geological Survey (Natural Environment Research Council). The authors publish with the permission of the Executive Director of the British Geological Survey.

References

Allen DJ, Brewerton LJ, Coleby LM, Gibbs BR, Lewis MA, MacDonald AM, Wagstaff SJ, Williams AT. 1997. The physical properties of major aquifers in England and Wales. In: British Geological Survey Technical Report (WD/97/34). Environment Agency R&D Publication 8, 312.

Bartl S, Schumberg S, Deutsch M. 2009. Revising time series of the Elberiver discharge for flood frequency determination at gauge Dresden. Natural Hazards and Earth System Sciences 9: 1805-1814.

Beven K. 2006. A manifesto for the equifinality thesis. Journal of Hydrology 320: 18-36. DOI:10.1016/j.jhydrol.2005.07.007

Beven K, Binley A. 1992. The future of distributed models - model calibration and uncertainty prediction. Hydrological Processes 6: 279-298. DOI:10.1002/hyp.3360060305

Beven K, Freer J. 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. Journal of Hydrology 249: 11 – 29. DOI:10.1016/s0022-1694(01)00421-8

Black AR, Law FM. 2004. Development and utilization of a national web-based chronology of hydrological events. Hydrological Sciences Journal 49: 237 – 246. DOI:10.1623/hysj.49.2.237.34835

Blaney HF, Criddle WD. 1950. Determining water requirements in irrigated areas from climatological and irrigation data. USDA Soil Conservation Service Tech. Paper No. 96: 48.

Bloomfield JP, Marchant BP. 2013. Analysis of groundwater drought building on the standardised precipitation index approach. Hydrology and Earth System Sciences 17: 4769 – 4787. DOI:10.5194/hess-17-4769-2013

Boorman DB, Hollis JM, Lilly A. 1995. Hydrology of soil types: a hydrologically-based classification of the soils of the United Kingdom. In: Report No. 126, Institute of Hydrology.

Breinholt A, Grum M, Madsen H, Thordarson FO, Mikkelsen PS. 2013. Informal uncertainty analysis (GLUE) of continuous flow simulation in a hybrid sewer system with infiltration in flow - consistency of containment ratios in calibration and validation? Hydrology and Earth System Sciences 17: 4159 – 4176. DOI:10.5194/hess-17-4159-2013

British Geological Survey. 2015. National Well Record Archive. www. bgs.ac.uk/research/groundwater/datainfo/NWRA.html (Accessed 23 February 2015)

Butterworth JA, Schulze RE, Simmonds LP, Moriarty P, Mugabe F. 1999. Hydrological processes and water resources management in a dryland environment IV: long-term groundwater level fluctuations due to variation in rainfall. Hydrology and Earth System Sciences 3: 353 – 361.

Calver A. 1997. Recharge response functions. Hydrology and Earth System Sciences 1: 47 – 53.

Chelmicki W, Ciszewski S, Zelazny M. 2002. Reconstructing groundwater level fluctuations in 20th century in the forested catchment of Drwinka (Niepolomice Forest, S. Poland). In UNESCO, Holko L, Miklanek P (eds); 203 – 208.

Chen ZH, Grasby SE. 2009. Impact of decadal and century-scale oscillations on hydroclimate trend analyses. Journal of Hydrology 365: 122 – 133. DOI:10.1016/j.jhydrol.2008.11.031

Chen ZH, Grasby SE, Osadetz KG. 2004. Relation between climate variability and groundwater levels in the upper carbonate aquifer, southern Manitoba, Canada. Journal of Hydrology 290: 43 – 62. DOI:10.1016/j.jhydrol.2003.11.029

Cole GA, Marsh TJ. 2006. An historical analysis of drought in England and Wales. In Climate Variability and Change - Hydrological Impacts, Demuth S, Gustard A, Planos E, Scatena F, Servat E (eds). Int Assoc Hydrological Sciences: Wallingford, UK; 483 – 489.

Conrads PA, Roehl EA. 2007. Hydrologic record extension of water-level data in the Everglades Depth Estimation Network (EDEN) using artificial neural network models, 2000 – 2006. In: U.S. Geological Survey Open-File Report 2007 – 1350.

Edwards DC, McKee TB. 1997. Characteristics of 20th century drought in the United States at multiple time scales. In: Climatology Report No. 97-2, Colorado State University.

Ferguson G, George SS. 2003. Historical and estimated ground water levels near Winnipeg, Canada, and their sensitivity to climatic variability. Journal of the American Water Resources Association 39: 1249 – 1259. DOI:10.1111/j.1752-1688.2003.tb03706.x

Field M. 1983. The Meteorological-Office rainfall and evaporation calculation system - MORECS. Agricultural Water Management 6: 297 – 306. DOI:10.1016/0378-3774(83)90017-3

Fowler HJ, Kilsby CG. 2002. A weather-type approach to analysing water resource drought in the Yorkshire region from 1881 to 1998. Journal of Hydrology 262: 177 – 192. DOI:10.1016/s0022-1694(02)00034-3

Franz KJ, Hogue TS. 2011. Evaluating uncertainty estimates in hydrologic models: borrowing measures from the forecast verification community. Hydrology and Earth System Sciences 15: 3367 – 3382. DOI:10.5194/hess-15-3367-2011

Freer J, Beven K, Ambroise B. 1996. Bayesian estimation of uncertainty in runoff prediction and the value of data: an application of the GLUE approach. Water Resources Research 32: 2161 – 2173. DOI:10.1029/95wr03723

Habets F, Gascoin S, Korkmaz S, Thiery D, Zribi M, Amraoui N, Carli M, Ducharne A, Leblois E, Ledoux E, Martin E, Noilhan J, Ottle C, Viennot P. 2010. Multi-model comparison of a major flood in the groundwater-fed basin of the Somme River (France). Hydrology and Earth System Sciences 14: 99 – 117.

Hanson RT, Dettinger MD, Newhouse MW. 2006. Relations between climatic variability and hydrologic time series from four alluvial basins across the southwestern United States. Hydrogeology Journal 14: 1122 – 1146. DOI:10.1007/s10040-006-0067-7

Holman IP, Rivas-Casado M, Bloomfield JP, Gurdak JJ. 2011. Identifying non-stationary groundwater level response to North Atlantic ocean-atmosphere teleconnection patterns using wavelet coherence. Hydrogeology Journal 19: 1269 – 1278. DOI:10.1007/s10040-011- 0755-9

Hughes AG, Vounaki T, Peach DW, Ireson AM, Jackson CR, Butler AP, Bloomfield JP, Finch J, Wheater HS. 2011. Flood risk from groundwater: examples from a Chalk catchment in southern England. Journal of Flood Risk Management 4: 143 – 155. DOI:10.1111/j.1753- 318X.2011.01095.x

Jackson CR, Bloomfield JP, Mackay JD. 2015. Evidence for changes in historic and future groundwater levels in the UK. Progress in Physical Geography 39: 49 – 67. DOI:10.1177/0309133314550668

Jenkins GJ, Perry MC, Prior MJ. 2008. The climate of the United Kingdom and recent trends. Met Office Hadley Centre: Exeter, UK.

Jones PD. 1984. River fl ow reconstruction from precipitation data. Journal of Climatology 4: 171 – 186.

Jones PD, Lister DH. 1998. River fl ow reconstructions for 15 catchments over England and Wales and an assessment of hydrologic drought since 1865. International Journal of Climatology 18: 999 – 1013. DOI:10.1002/(sici)1097-0088(199807)18:9 < 999::aid-joc300 > 3.0.co;2-8

Jones PD, Lister DH, Wilby RL, Kostopoulou E. 2006. Extended river flow reconstructions for England and Wales, 1865 – 2002. International Journal of Climatology 26: 219 – 231. DOI:10.1002/joc.1252

Jousma G, Roelofsen FJ. 2004. World-wide inventory on groundwater monitoring International Groundwater Resources Assessment Centre. Utrecht, Netherlands.

Kay AL, Bell VA, Blyth EM, Crooks SM, Davies HN, Reynard NS. 2013. A hydrological perspective on evaporation: historical trends and future projections in Britain. Journal of Water and Climate Change 4: 193 – 208. DOI:10.2166/wcc.2013.014

Legg TP. 2011. Determining the accuracy of gridded climate data and how this varies with observing-network density. Advances in Science and Research 6: 195 – 198. DOI:10.5194/asr-6-195-2011

Li HB, Sheffield J, Wood EF. 2010a. Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching. Journal of Geophysical Research-Atmospheres 115: DOI:10.1029/2009jd012882

Li L, Xia J, Xu CY, Singh VP. 2010b. Evaluation of the subjective factors of the GLUE method and comparison with the formal Bayesian method in uncertainty assessment of hydrological models. Journal of Hydrology 390: 210 – 221. DOI:10.1016/j.jhydrol.2010.06.044

Little MA, Bloomfield JP. 2010. Robust evidence for random fractal scaling of groundwater levels in unconfined aquifers. Journal of Hydrology 393: 362 – 369. DOI:10.1016/j.jhydrol.2010.08.031

Lloyd-Hughes B, Prudhomme C, Hannaford J, Parry S, Keef C, Rees G. 2009. The Spatial Coherence of European Droughts – UK and European Drought catalogues. Environment Agency: Bristol, UK.

Mackay JD, Jackson CR, Wang L. 2014. A lumped conceptual model to simulate groundwater level time-series. Environmental Modelling & Software 61: 229 – 245. DOI:10.1016/j.envsoft.2014.06.003

Marsh T, Hannaford J. 2008. UK Hydrometric Register Hydrological data UK series. Centre for Ecology and Hydrology, NERC: Wallingford, UK.

Marsh T, Cole G, Wilby R. 2007. Major droughts in England and Wales, 1800 – 2006. Weather 62: 87 – 93. DOI:10.1002/wea.67

McKee TB, Doesken NJ, Leist J. 1993. The relationship of drought frequency and duration to time scales. In: 8th Conference on Applied Climatology.

Met Office. 2014. UK rainfall, sunshine and temperature time-series. UK Meteorological Office. www.metof fi ce.gov.uk/climate/uk/actualmonthly/ (Accessed 23 February 2015)

Misstear BDR, Beeson S. 2000. Using operational data to estimate the reliable yields of water-supply wells. Hydrogeology Journal 8: 177–187. DOI:10.1007/s100400050004

Mittman T, Band LE, Hwang T, Smith ML. 2012. Distributed hydrologic modeling in the Suburban landscape: assessing parameter transferability from gauged reference catchments. Journal of the American Water Resources Association 48: 546 – 557. DOI:10.1111/j.1752-1688.2011.00636.x

Mo XG, Pappenberger F, Beven K, Liu SX, De Roo A, Lin ZH. 2006. Parameter conditioning and prediction uncertainties of the LISFLOOD-WB distributed hydrological model. Hydrological Sciences Journal 51: 45 – 65. DOI:10.1623/hysj.51.1.45

Monteith JL. 1965. Evaporation and environment. 19th Symposia of the Society for Experimental Biology, University Press Cambridge, 205 – 234.

Monteith JL, Unsworth MH. 2008. Principles of Environmental Physics. Elsevier: Oxford, UK.

Najib K, Jourde H, Pistre S. 2008. A methodology for extreme groundwater surge predetermination in carbonate aquifers: Groundwater flood frequency analysis. Journal of Hydrology 352: 1 – 15. DOI:10.1016/j.jhydrol.2007.11.035

Nash JE, Sutcliffe JV. 1970. River fl ow forecasting through conceptual models part I - a discussion of principles. Journal of Hydrology 10: 282 – 290.

Oudin L, Hervieu F, Michel C, Perrin C, Andreassian V, Anctil F, Loumagne C. 2005. Which potential evapotranspiration input for a lumped rainfall-runoff model? Part 2 - towards a simple and efficient potential evapotranspiration model for rainfall-runoff modelling. Journal of Hydrology 303: 290 – 306. DOI:10.1016/j.jhydrol.2004.08.026

Pereira LS, Perrier A, Allen RG, Alves I. 1999. Evapotranspiration: concepts and future trends. Journal of Irrigation and Drainage Engineering-ASCE 125: 45 – 51. DOI:10.1061/(asce)0733-9437(1999) 125:2(45)

Perez-Valdivia C, Sauchyn D. 2011. Tree-ring reconstruction of groundwater levels in Alberta, Canada: Long term hydroclimatic variability. Dendrochronologia 29: 41 – 47. DOI:10.1016/j.dendro.2010.09.001

Perry M, Hollis D. 2005. The generation of monthly gridded datasets for a range of climatic variables over the UK. International Journal of Climatology 25: 1041 – 1054. DOI:10.1002/joc.1161

Peters E, Bier G, van Lanen HAJ, Torfs P. 2006. Propagation and spatial distribution of drought in a groundwater catchment. Journal of Hydrology 321: 257 – 275. DOI:10.1016/j.jhydrol.2005.08.004

Phillips ID, McGregor GR. 1998. The utility of a drought index for assessing the drought hazard in Devon and Cornwall, South West England. Meteorological Applications 5: 359 – 372.

Piñol J, Beven K, Viegas D. 2005. Modelling the effect of fi re-exclusion and prescribed fire on wild fi re size in Mediterranean ecosystems. Ecological Modelling 183: 397 – 409. DOI:10.1016/j.ecolmodel.2004.09.001

Price K, Thomas Purucker S, Kraemer SR, Babendreier JE, Knightes CD. 2014. Comparison of radar and gauge precipitation data in watershed models across varying spatial and temporal scales. Hydrological Processes 9: 3505 – 3520. DOI:10.1002/hyp.9890

Prudhomme C, Haxton T, Crooks S, Jackson C, Barkwith A, Williamson J, Kelvin J, Mackay J, Wang L, Young A, Watts G. 2013. Future Flows Hydrology: an ensemble of daily river flow and monthly groundwater levels for use for climate change impact assessment across Great Britain. Earth System Science Data 5: 101 – 107. DOI:10.5194/essd-5-101-2013

Risk Management Solutions. 2007. 1947 UK river floods: 60-year retrospective. In: RMS Special Report.

Royal Meteorological Society. 1948. The weather of 1947 in Great Britain. Weather 3: 27 – 30. DOI:10.1002/j.1477-8696.1948.tb00856.x

Rushton KR. 2003. Groundwater hydrology: conceptual and computational models. John Wiley and Sons Ltd: Chichester, UK.

Rutulis M. 1989. Groundwater drought sensitivity of Southern Manitoba. Canadian Water Resources Journal 14: 18 – 33. DOI:10.4296/cwrj1401018

Sapriza-Azuri G, Jódar J, Navarro V, Jan Slooten L, Carrera J, Gupta HV. 2015. Impacts of rainfall spatial variability on hydrogeological response. Water Resources Research 51: 1300 – 1314. DOI:10.1002/2014WR016168

Schilling W, Einfalt T. 1982. A multivariate stochastic-model for the reconstruction of groundwater data. Hydrological Sciences Journal 27: 224 – 225.

Shen ZY, Chen L, Chen T. 2012. Analysis of parameter uncertainty in hydrological and sediment modeling using GLUE method: a case study of SWAT model applied to Three Gorges Reservoir Region, China. Hydrology and Earth System Sciences 16: 121 – 132. DOI:10.5194/hess-16-121-2012

Stoll S, Franssen HJH, Barthel R, Kinzelbach W. 2011. What can we learn from long-term groundwater data to improve climate change impact studies? Hydrology and Earth System Sciences 15: 3861 – 3875. DOI:10.5194/hess-15-3861-2011

Viola F, Noto LV, Cannarozzo M, La Loggia G. 2009. Daily stream flow prediction with uncertainty in ephemeral catchments using the GLUE methodology. Physics and Chemistry of the Earth 34: 701 – 706. DOI:10.1016/j.pce.2009.06.006

Wen L. 2009. Reconstruction natural fl ow in a regulated system, the Murrumbidgee River, Australia, using time series analysis. Journal of Hydrology 364: 216 – 226. DOI:10.1016/j.jhydrol.2008.10.023

Williams A, Bloomfield J, Griffiths K, Butler A. 2006. Characterising the vertical variations in hydraulic conductivity within the Chalk aquifer. Journal of Hydrology 330: 53 – 62. DOI:10.1016/j.jhydrol.2006.04.036

Xiong L, O ’ Connor KM. 2008. An empirical method to improve the prediction limits of the GLUE methodology in rainfall-runoff modeling. Journal of Hydrology 349: 115 – 124. DOI:10.1016/j.jhydrol.2007.10.029

Tables

Table I. Observation borehole details and associated aquifer property information after Allen et al. (1997)

Observation Borehole Name	Aquifer	Confinement	Borehole Depth (m)	Transmissivity (m^2day^{-1})		Estimated specific yield (%)
				Median	Interquartile range	
Bussels No.7A	Sandstone	Unconfined	91	105	30-303	5-20
Chilgrove House	Chalk	Unconfined	62	440	230-1600	0.5-2
Lower Barn Cottage	Lower Greensand	Unconfined	~9 m	270	140-500	10-20
New Red Lion	Limestone	Varies with water level	50	660	259-2265	0.5-5
Skirwith	Sandstone	Possibly confined at borehole, unconfined regionally	89	Limited data: typically 10s-100s		5-15
Swan House	Limestone	Confined	60	205	139-564	0.5-5

Table II. AquiMod model parameters

Model component	Parameter (units)	Description	Bussels No.7	Chilgrove House	Lower Barn Cottage	New Red Lion	Skirwith	Swan House
Soil	BFI (-)	Catchment baseflow index	0.54	0.81	0.8	0.49	0.43	0.39
	FC (-)	Field capacity of the soil	0.18	0.29	0.29	0.29	0.29	0.29
	WP (-)	Wilting point of the soil	0.09	0.15	0.19	0.19	0.19	0.19
	Zr (mm)	Maximum rooting depth of vegetation	100-3000	100-3000	100-3000	326-1911	100-3000	253-2569
	p (-)	Depletion factor of vegetation	0.01-0.99	0.1-0.99	0.1-0.99	0.01-0.99	0.01-0.99	0.01-0.99
Unsaturated Zone	n (-)	Maximum number of time-steps taken for soil drainage to reach groundwater	5	5	7	5	6	7
	k (-)	Weibull shape parameter	1-7	1-7	1-7	1-7	1-7	1-7
	λ (-)	Weibull scale parameter	1-3	1-3	1-3	1-3	1-3	2.5-5
Saturated Zone	K_3 (m d^{-1})	Top layer hydraulic conductivity	10-150	3.5-100	10-100	1-100	50-200	2.9-80
	K_2 (m d^{-1})	Middle layer hydraulic conductivity	0.001-0.01	0.1-10	1-23	0.1-10	0.01-0.1	0.1-4.6
	K_1 (m d^{-1})	Bottom layer hydraulic conductivity	0.001-0.01	0.01-1	0.01-0.5	0.01-1	0.01-0.07	0.01-1
	S (%)	Aquifer storage coefficient	5-20	0.5-2.3	5-20	0.5-1.4	5-20	0.5-2.2
	z_3 (masl)	Top outlet elevation	22.9-24.4	33.5-77.2	10.1-13.6	3.3-23.7	129.4-130.3	77.7-89.7
	z_2 (masl)	Middle outlet elevation	-25.7	27.2	8.0	-5.4	82.2	60.3
	z_1 (masl)	Bottom outlet elevation	-74.3	-11.4	3.0	-14.2	35.0	43.0
	Δx (m)	Catchment length	420	3000	500	4000	1300	4000

Table III. Summary of model performance

Site	Calibration Period	Number of simulations	Number of behavioural models	Maximum NSE	Evaluation Period	Number of behavioural models	Maximum NSE	Reconstruction Period
Bussels No.7A	Jan-1972 to Aug-1991	10^6	39,077	0.93	Sep-1991 to Mar-2012	12,416	0.88	Jan-1910 to Dec-1971
Chilgrove House	Jan-1990 to Mar-2012	10^6	93,079	0.87	Jan-1970 to Dec-1989	53,075	0.93	Jan-1910 to Dec-1969
Lower Barn Cottage	Jan-1998 to Mar-2012	10^6	62,055	0.92	May-1975 to Dec-1997	16,246	0.87	Jan-1910 to Apr-1975
New Red Lion	Apr-1964 to Feb-1987	10^7	11,533	0.77	Mar-1987 to Mar-2012	2,940	0.70	
Skirwith	Aug-1993 to Mar-2012	10^6	33,255	0.91	Nov-1978 to Jul-1993	12,669	0.84	Jan-1910 to Oct-1978
Swan House	Sep-1973 to Jun-1991	10^7	1,885	0.73	Jul-1991 to Feb-2010	653	0.76	

Table IV. Containment ratios (%) of 5 and 95% likelihood-weighted prediction limits based on an NSE threshold of 0.5

	Calibration	Evaluation		
	All levels	All levels	Levels < median	Levels > median
Bussels No.7A	94.5	83.7	86.1	81.1
Chilgrove House	85.7	86.6	100	73.1
Lower Barn Cottage	90.0	81.5	81.5	81.5
New Red Lion	59.5	64.0	45.3	82.7
Skirwith	90.3	79.5	77.3	81.8
Swan House	67.1	66.8	69.4	64.9

Table V. Summary of 20th century drought events in England

Year	Comments	Reference
1921-2	Major drought which was very severe across much of England and Wales including East Anglia and the South-East. Episodic in north-west England.	Marsh et al., 2007
1933-4	Major drought which was intense across southern Britain.	Marsh et al., 2007
1943-4	Severe drought during which flows in many rivers across the English Lowlands were exceptionally low.	Marsh et al., 2007
1947	Ranked 14th in severity of water resource drought for Yorkshire region over period 1881-1998.	Fowler and Kilsby, 2002; Royal Meteorological Society, 1948
1949	Ranked 13th in severity of water resource drought for Yorkshire region over period 1881-1998.	Fowler and Kilsby, 2002
1959	Three-season drought, which was most severe in eastern, central and north-eastern England. Modest groundwater impact.	Marsh et al., 2007
1962-4	Some long droughts (1962 – 1964; 1995 – 1997; 1988 – 1992) result from a combination of both winter and summer deficiencies (Lloyds-Hughes et al., 2009). The 1964/5 drought over south-west England had two sequences of core months: January-April 1964 and September 1964-April 1965 (Phillips and McGregor, 1998).	Lloyds-Hughes et al., 2009; Phillips and McGregor, 1998
1976	UK benchmark drought. Sever impacts on groundwater and river flows across UK.	Marsh et al., 2007
1990-2	Major drought causing exceptionally low groundwater levels in summer 1992.	Marsh et al., 2007
1995-7	Long duration drought with three intense episodes. Very low groundwater levels particularly during summer 1995.	Marsh et al., 2007

Table VI. Chilgrove House groundwater levels at selected percentiles of the observed and reconstructed distributions for the period

Exceedance probability	Observed	Reconstructed	Reconstructed (bias corrected rainfall)
10	64.9	65.8	60.0
25	55.6	57.7	52.2
50	46.8	49.5	45.2
75	42.0	43.0	40.5
90	39.0	39.7	37.7

Figures

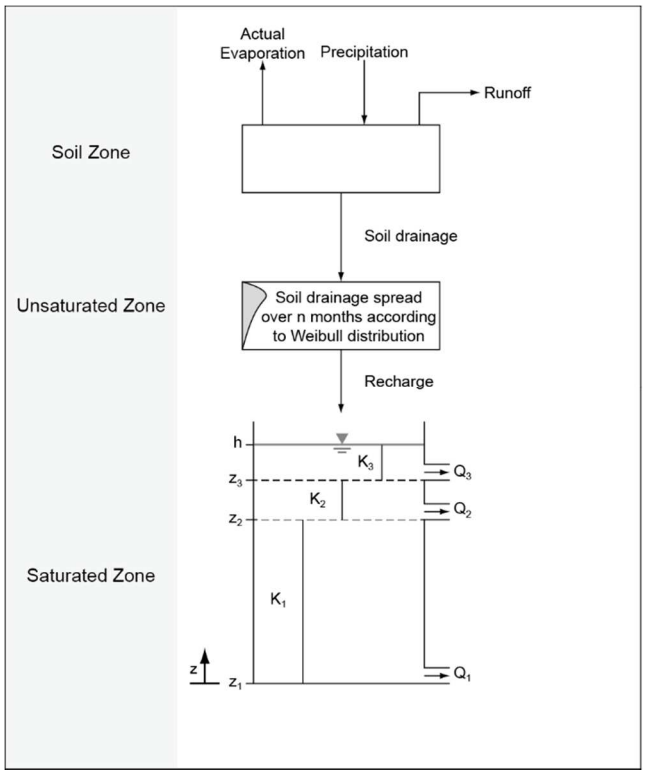

Figure 1. AquiMod model structure

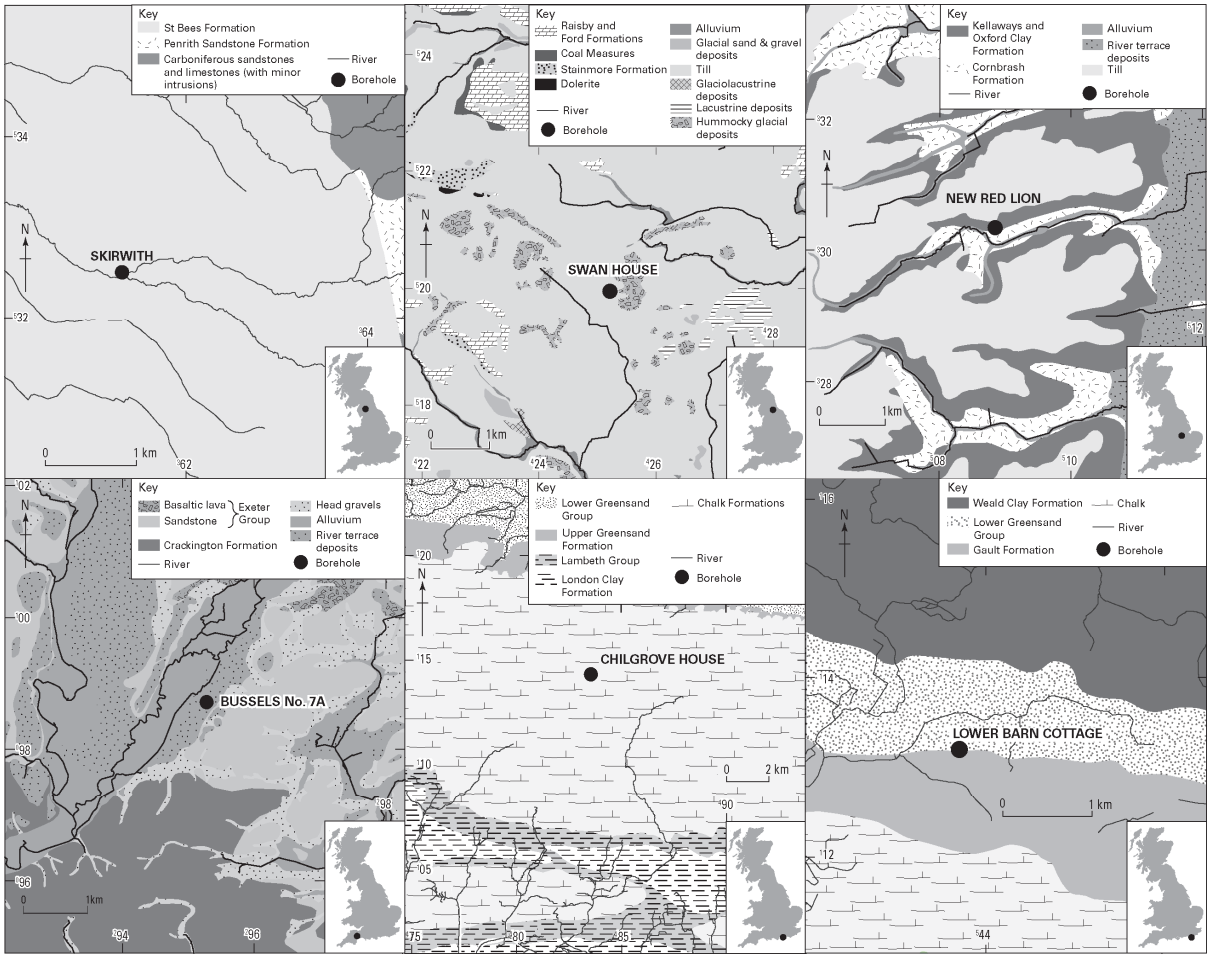

Figure 2. Observation borehole locations. Contains Ordnance Survey data © Crown Copyright and database rights [2016]

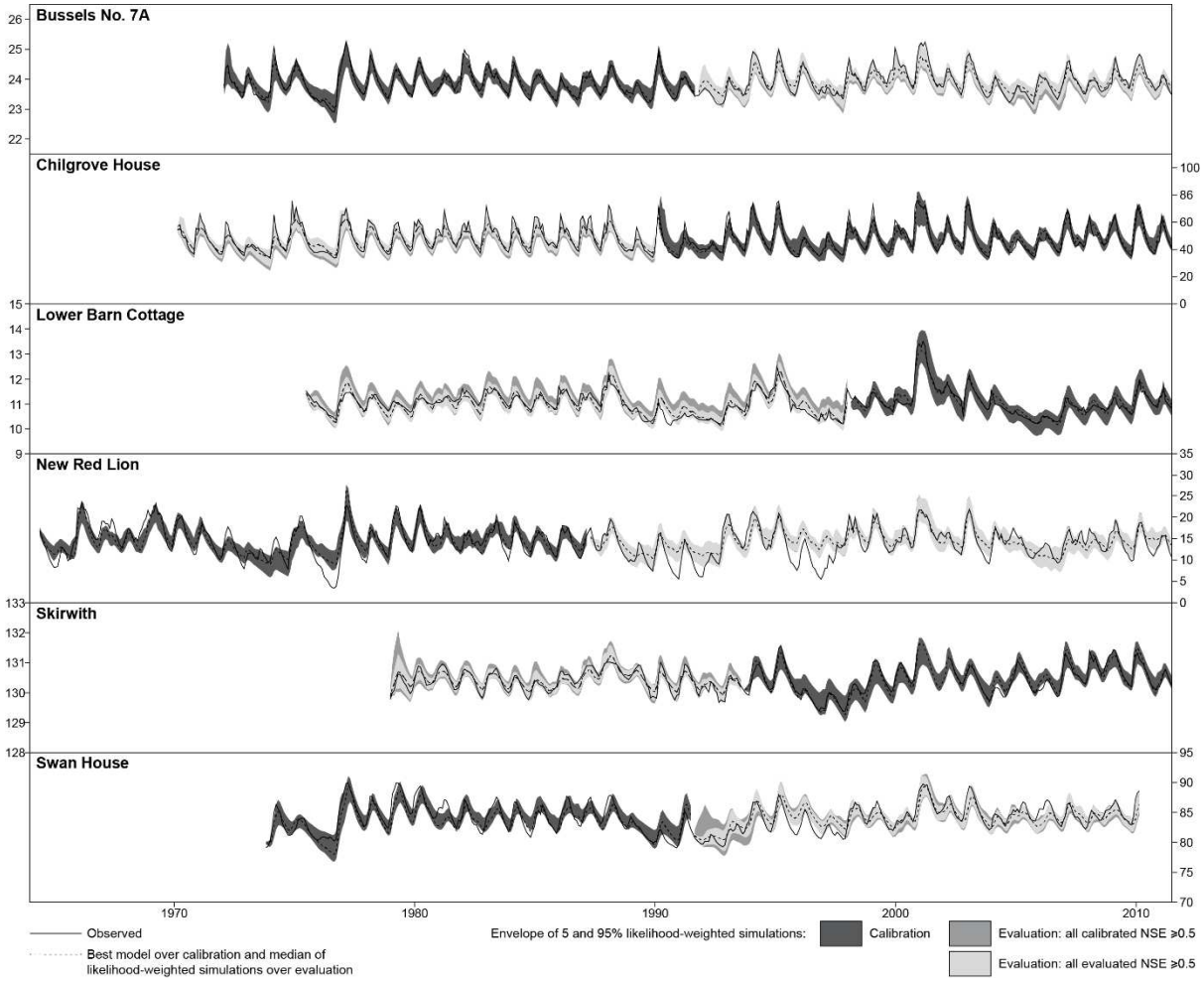

Figure 3. Observed and simulated groundwater level time-series over calibration and evaluation periods

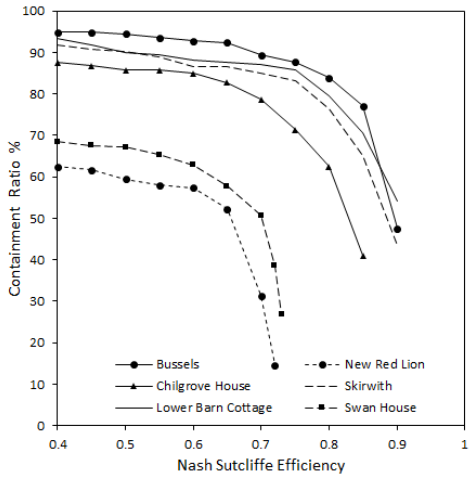

Figure 4. Containment ratios of 5 and 95% likelihood-weighted prediction limits during calibration periods for varying thresholds of NSE applied in GLUE

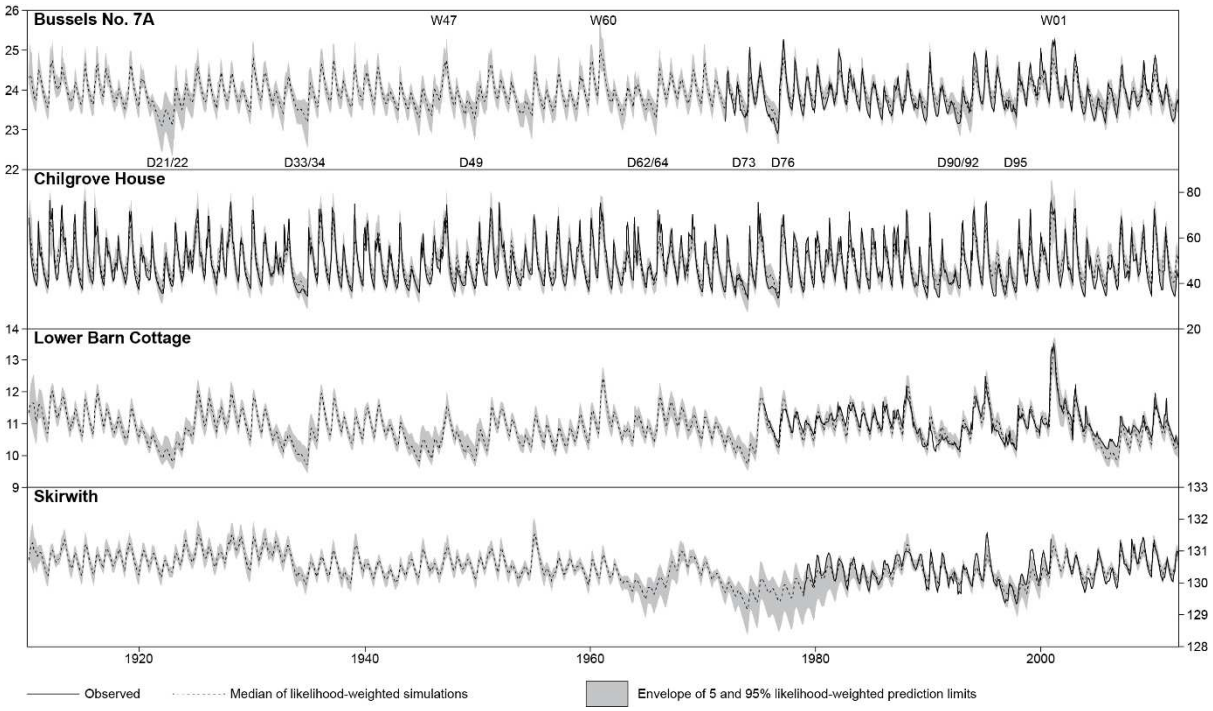

Figure 5. Reconstructed and observed groundwater level time-series. Notable drought (D, year) and wet (W, year) periods are identified on the upper panel

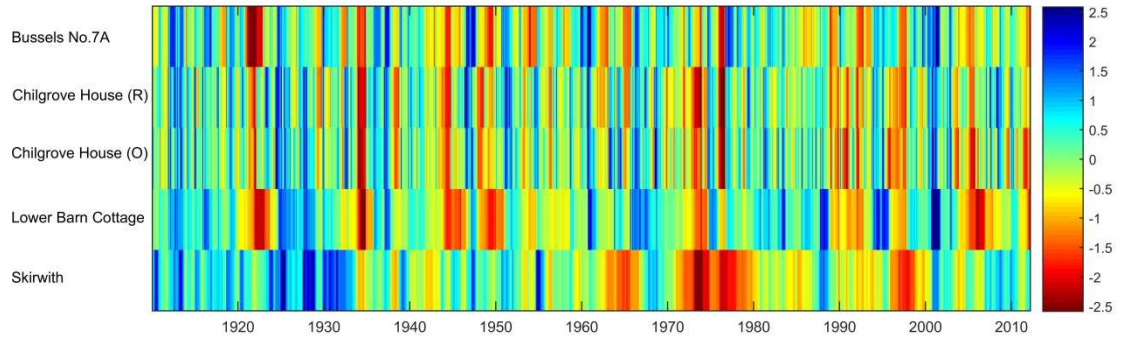

Figure 6. Standard Groundwater level Index (SGI) heat map based on best calibrated models. For Chilgrove House both the reconstructed (R) and (O) series are plotted

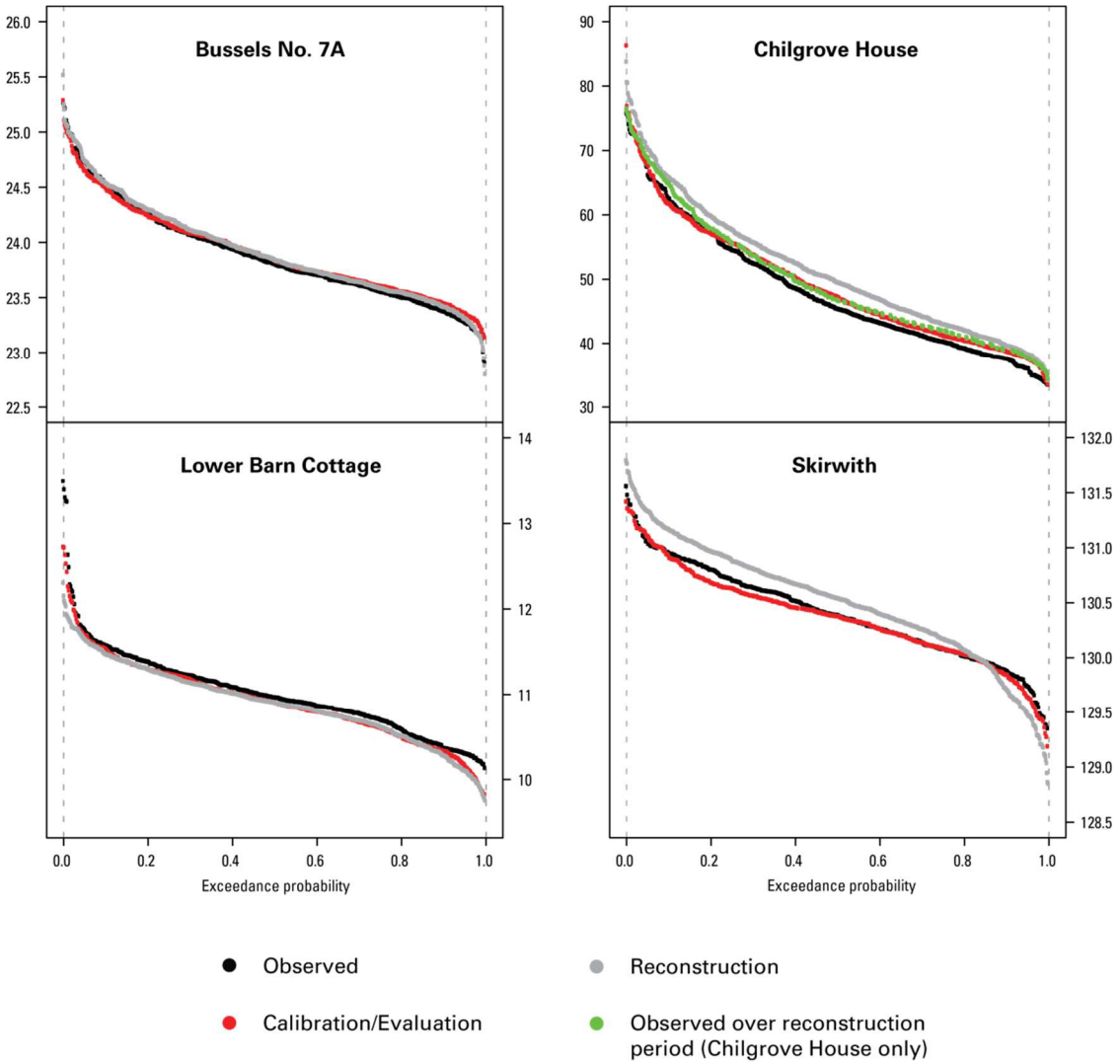

Figure 7. Simulated groundwater level (m a.s.l.) duration curves

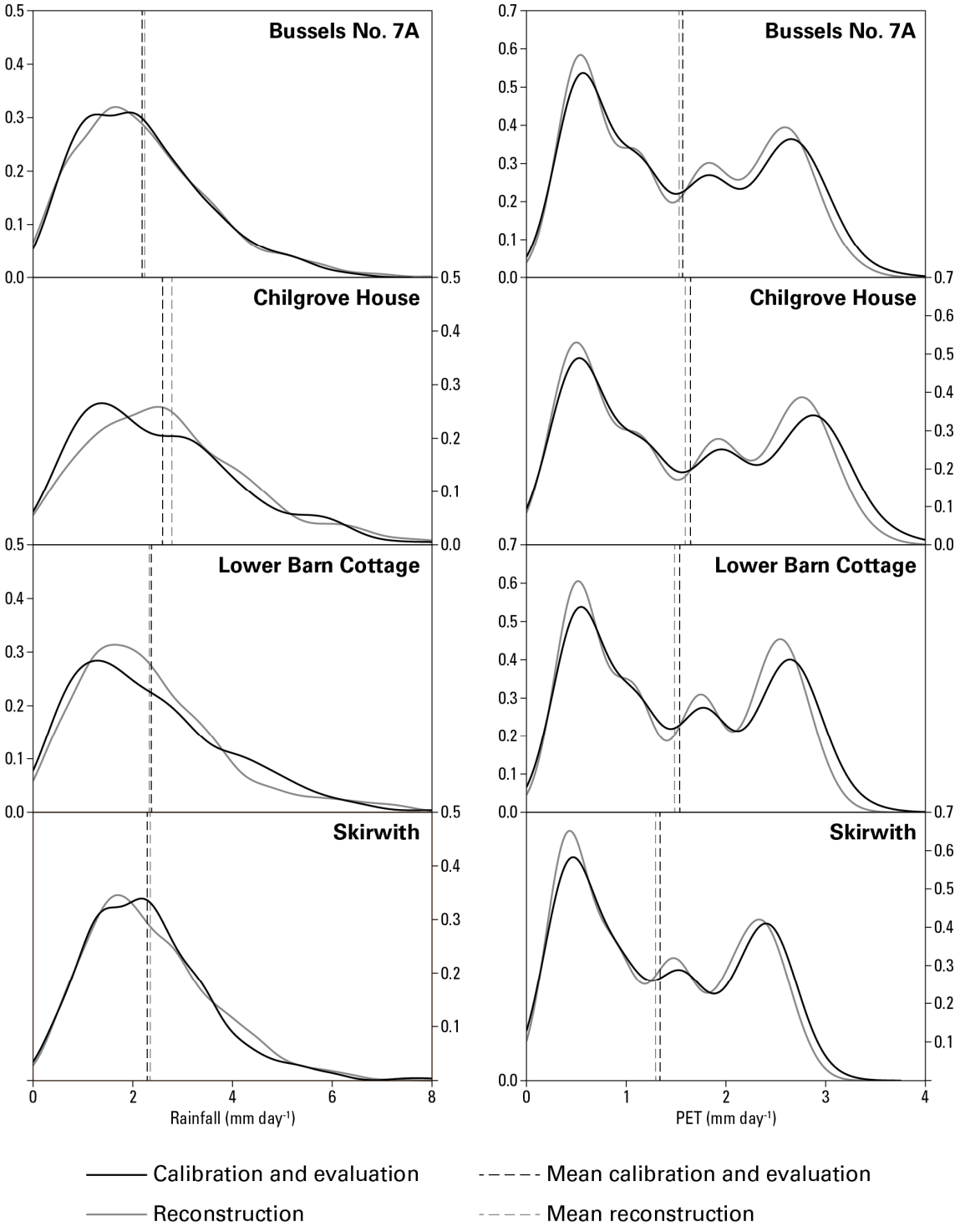

Figure 8. Kernel density plots of monthly rainfall and PET

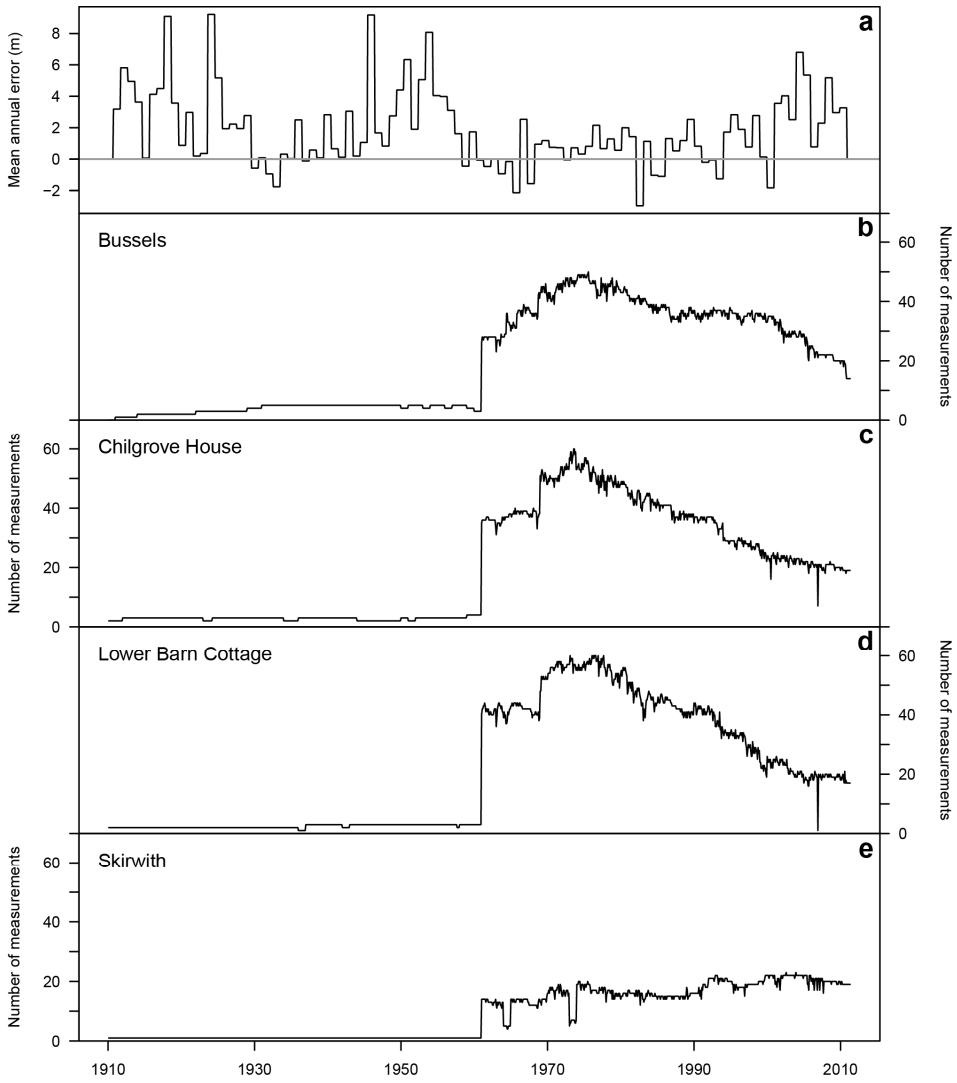

Figure 9. Mean annual error of the Chilgrove House model over the full simulation period (a), and time-series of the average number of rainfall measurements made each month within a 20 km radius of the four reconstructed sites (b – e)