TITLE:

A lumped conceptual model to simulate groundwater level time-series

AUTHORS:

J.D. Mackay, C.R. Jackson*, L. Wang

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

* Corresponding author: Tel.: +44 1159 363100; fax: +44 1159 363200; Email: crja@bgs.ac.uk

ABSTRACT:

Lumped, conceptual groundwater models can be used to simulate groundwater level time-series quickly and efficiently without the need for comprehensive modelling expertise. A new model of this type, AquiMod, is presented for simulating groundwater level time-series in unconfined aquifers. Its modular design enables users to implement different model structures to gain understanding about controls on aquifer storage and discharge. Five model structures are evaluated for four contrasting aquifers in the United Kingdom. The ability of different model structures and parameterisations to replicate the observed hydrographs is examined. AquiMod simulates the quasi-sinusoidal hydrographs of the relatively uniform Chalk and Sandstone aquifers most efficiently. It is least efficient at capturing the flashy hydrograph of a heterogeneous, fractured Limestone aquifer. The majority of model parameters demonstrate sensitivity and can be related to available field data. The model structure experiments demonstrate the need to represent vertical aquifer heterogeneity to capture the storage-discharge dynamics efficiently.

KEYWORDS:

Groundwater level simulation; Observation borehole hydrograph; Lumped conceptual modelling; AquiMod

SOFTWARE:

Name of software: AquiMod

Description: A lumped modular conceptual groundwater model for simulating groundwater level time-series at observation boreholes

Developers: Jonathan Mackay (joncka@bgs.ac.uk) and Christopher Jackson (crja@bgs.ac.uk)

Program language: C++

Availability: On request under licence.

1. Introduction

Groundwater aquifers are complex, non-linear and heterogeneous systems that respond to natural and human influences including climate, land use and groundwater abstractions (Taylor and Alley, 2001). Our capacity to measure these hydrogeological complexities in the field is limited. Currently, regular and long-term groundwater level measurements taken from observation boreholes provide the best indication of an aquifer's flow and storage behaviour and allow us to differentiate between natural and anthropogenic stresses on groundwater. Thus extensive, accurate and continuous groundwater level time-series data are fundamental to understanding and managing our groundwater resources effectively (Alley et al., 2002).

Continuous groundwater level time-series contain information on the seasonality and trend in groundwater levels. With sufficient data, they also provide information on the frequency and magnitude of extreme events which can then be used to infer patterns of groundwater drought for example (Bloomfield and Marchant, 2013). Access to this information is essential as rates of groundwater abstraction increase with demand and potential stresses induced by climate change and urbanisation may take effect (Wada et al., 2010).

Measurements from pumping tests and one-off dip readings comprise the majority of groundwater level monitoring datasets which typically run for a period of days or weeks and are unsuitable for investigating extreme events or long-term trends (Taylor and Alley, 2001). Computational models provide an alternative means of obtaining groundwater level datasets through simulation rather than observation and can also be used to derive aquifer hydrogeological properties (Peck, 1988). These tools have been used in the past to

reconstruct and extend groundwater level records (Conrads and Roehl, 2007), forecast extreme events in the near future (Adams et al., 2008; Daliakopoulos et al., 2005) as well as aid investigations into long-term water resource sustainability under climate and land use stresses (Goderniaux et al., 2009; Jackson et al., 2011; Sun et al., 2011).

A diverse range of groundwater modelling approaches have been applied in the past to simulate groundwater levels. Physically based, process-driven models remain the most widely used. They are based on simplifications of physical laws of fluid dynamics to simulate flow through the subsurface. The equations that they employ are often complex, but offer the advantage that they use parameters that it may be possible to relate to known hydrogeological properties. The equations are typically approximated numerically across a multi-dimensional spatial grid and through time using, for example, finite difference methods. This type of distributed modelling makes it feasible to include complex heterogeneity and anisotropy. Shepley et al. (2012) describe the benefits of applying distributed numerical groundwater models for understanding complex groundwater flow systems and assessing the impact of human and environmental stresses on groundwater resources. Indeed, distributed models have been applied to some of the world's major aquifers for these purposes (Gossel et al., 2004; Scanlon et al., 2003; Smith and Welsh, 2011).

An alternative is to disregard physical theories of groundwater flow and instead, derive an entirely empirical relationship between groundwater levels and one or a more predictor variables. A number of studies have been undertaken in which such empirical approaches have been applied to simulate groundwater level time-series. Typically these have used classical 'black box' modelling methodologies (Jakeman et al., 2006), such as statistical transfer function models or neural networks. For example the Box-Jenkins autoregressive integrated moving average (ARIMA) model has been used in a number of studies to simulate meteorological impacts on groundwater levels (Aflatooni and Mardaneh, 2011; Ahn, 2000; Gemitzi and Stefanopoulos, 2011) as it can incorporate the complex seasonal, non-stationary and random components observed in groundwater level time-series. Simpler empirical regression methods have also been employed including Bloomfield et al. (2003)

who constructed multiple linear regression models for several UK catchments to simulate annual minimum groundwater level time-series from rainfall datasets from which they could explain between 50-84% of the variance in historic levels. Other non-parametric methods have been used, perhaps most extensively, artificial neural networks (ANN), which have been shown to be very proficient at pattern recognition in time-series datasets. They have been applied successfully to aquifers in both arid (Coulibaly et al., 2001) and tropical regions (Ghose et al., 2010)**,** karstic aquifers (Trichakis et al., 2011) and for forecasting groundwater levels with acceptable predictions up to 18 months ahead (Daliakopoulos et al., 2005; Sreekanth et al., 2009)**.** Maier and Dandy (2000) and Dawson and Wilby (2001) provide comprehensive reviews on the use of ANNs for hydrological modelling.

Whilst black-box modelling methodologies can be useful, it is recognised that they generally provide little insight into the controls on the behaviour of a system (Lees, 2000; Young et al., 2007). Consequently, Young and Lees (1993) propounded a more structured strategy for the development and application of environmental simulation models, namely 'data-based mechanistic' (DBM) modelling. In this approach prior assumptions about the form of the appropriate model to apply are minimised to avoid the incorporation of prejudicial perceptions about the structure of the model that is required. In fact, the DBM approach incorporates a number of stages, which are aimed at maximising insight into the behaviour of the system being studied. Specifically, for example, it seeks to convert deterministic simulations into stochastic forms, through, for example, the use of Monte Carlo simulations, and to identify parsimonious models through a process of model simplification. Whilst this is a much more rigorous approach to model development and simulation it is probably less widely used because it is generally a more involved and time-consuming process. For example, it is possible that the simulation of groundwater level time-series at a number of different sites would require the identification of a number of different models structures within the DBM process. To the best of our knowledge no examples of the application of the DBM approach to simulate groundwater levels have been reported in the peer-reviewed literature. This is probably partly because the DBM approach has been developed within the hydrological modelling research community but also because of the dominance of the use of distributed models in hydrogeological studies. An example of the application of the

DBM approach to the simulation of levels rather than flows is provided by Romanowicz et al. (2006) who simulated in-channel water levels at various locations within the River Severn catchment in the UK.

Similarly to black-box models, physically-based models have their disadvantages. While physically-based models allow the inclusion of complex non-linear processes that are thought to exist in aquifer systems, characterizing these complexities requires a detailed conceptual understanding of the system, numerous model parameters and more hydrogeological and climate data than are typically available, making this type of modelling inherently uncertain (Beven, 2001; Konikow and Bredehoeft, 1992). Furthermore, physically based models are often costly and time consuming to build and demand highly skilled and proficient modellers to operate them, rendering them inaccessible to many hydrogeologists.

Conceptual, lumped parameter modelling is an alternative approach that neglects some of the complexities incorporated in physically-based models, but maintains some fundamental physical principles from our conceptual understanding of groundwater systems. These types of models can be assessed against and constrained by physically measured field data, but are simple enough to be run quickly, at little cost and without the need for broad modelling expertise. They also require fewer parameters than physical models, making them easier to constrain through automated calibration procedures. Birtles and Reeves (1977) developed one of the earliest lumped parameter groundwater models where the groundwater system was generalised as four units to represent unconfined and confined groundwater zones, a superficial layer and a spring unit. All units were lumped in the horizontal and vertical, but could interact through vertical and lateral fluxes. They then used the model to derive optimum abstraction regimes for multiple water-supply boreholes. Indeed others have emulated this approach since (Anaya and Wanakule, 1993; Barrett and Charbeneau, 1997; Kazumba et al., 2008; Thiéry, 2012). Lumped models have also been used to simulate spring flows and groundwater levels in highly heterogeneous aquifers (Keating, 1982), estimate aquifer parameters (Olin, 1995), and to develop early warning systems for groundwater flooding (Adams et al., 2008). Flores W et al. (1978) and Pozdniakov and Shestakov (1998) made stochastic simulations of groundwater levels using linear reservoirs with parameters

drawn from a probability distribution. Flores W et al. (1978) go further and combine their model with a simple economic model to calculate optimum pumping borehole operating policies. Barrett and Charbeneau (1997) demonstrated that groundwater level simulations from lumped conceptual models can be comparable to physical models when they applied both to the Edwards aquifer, USA. Eberts et al. (2012) compared the results from a 3D physically-based groundwater particle tracking model (Pollock, 1994) and a number of lumped parameter models. Specifically, they compared the simulated age distributions of groundwater at a number of wells in four contrasting aquifers in the US as a means to quantify contamination vulnerability and found that they gave remarkably similar results.

None of these studies that have used lumped parameter models to simulate groundwater levels explicitly investigated the appropriateness of the form of the relationship between groundwater level, storage and discharge adopted within their model structures. Rather, they each applied a deterministic representation of this relationship justified based on conceptual understanding. Moore and Bell (2002) briefly discuss what form this function should take, and present the appropriate type of linear or non-linear store for different aquifer types, based on a consideration of the Horton-Izzard model (Dooge, 1973) and standard groundwater theory (Todd, 1959). They incorporate a representation of groundwater discharge into their PDM rainfall-runoff model, which is then used to simulate flow in the River Lavant, and the groundwater level in an observation borehole in this Chalk catchment in south-east England. The suitability of a linear form for homogeneous constant transmissivity aquifers, and of a quadratic form for homogeneous unconfined aquifers, is summarised. However, they also state that cubic forms have been found to be useful in practical applications of the PDM model and then adopt this type of store in the model that they apply to simulate the groundwater level time-series of their study catchment borehole, West Dean Nursery. The simulated groundwater hydrograph is reasonable but their model is poor at reproducing groundwater recession and the observed variability in the annual minima. This is likely to be because the model cannot capture the heterogeneity in the hydraulic conductivity and storage structure of the Chalk aquifer. By contrast Keating (1982) adopted a conceptualisation of the vertical variation of hydraulic conductivity with depth of the Chalk informed by the hydrogeology of this aquifer system (Allen et al., 1997;

Bloomfield, 1997; Williams et al., 2006). In this model, a classical 'cocktail glass' representation of the Chalk depth-hydraulic conductivity profile (Rushton, 2003) was used to describe the increase in hydraulic conductivity of the aquifer in the zone of water table fluctuation.

In summary, computational groundwater models have been used extensively to simulate groundwater level time-series for a variety of applications. A wide range of modelling approaches exist, the choice of which is likely to reflect the user's background and expertise in modelling, the availability of field data, and the purpose of the modelling exercise. Simple, lumped conceptual models have been shown to be an efficient means to simulate groundwater level time-series. They include some physical principles from our conceptual understanding of groundwater systems which means that their structure and parameters can be assessed against and constrained by measured field data. They also employ simple algorithms and can be run quickly and efficiently without the need for broad modelling expertise. Of course, the suitability of applying simple, lumped process representations to systems that are known to be highly non-linear and heterogeneous should always be considered. As such, it is important to test the appropriateness of different representations of the relationship between groundwater storage and discharge. In particular, the incorporation of vertically heterogeneous properties of hydraulic conductivity in simple lumped models has been neglected even though it has often been shown to be very important for capturing groundwater storage-discharge dynamics (Keating, 1982; Rushton and Rathod, 1981; Rushton et al., 1982).

In this study, we present a new modular, modelling framework, AquiMod, which has been developed to simulate groundwater level time-series at observation boreholes in unconfined aquifers by linking simple conceptual hydrological algorithms that represent soil drainage, the transfer of water through the unsaturated zone and groundwater flow. These algorithms are based on established hydrological concepts and employ parameters that can be assessed against field data. AquiMod represents a significant development over previous lumped parameter models because it has been designed to include multiple representations of groundwater flow and also vertically heterogeneous hydraulic conductivity parameters.

In this study, five different model structures are applied to four contrasting aquifers known to possess complex hydrogeological characteristics. This research aims to determine first, if AquiMod is able to simulate groundwater levels efficiently in these different settings. Second, if the optimised model parameters are related to, and thus can be constrained by, available field data. Finally, by analysing the impact of using gradually simpler representations of groundwater flow on the model behaviour at each study site, this paper explores the importance of including vertically heterogeneous aquifer properties and uses this analysis to gain understanding about storage-discharge controls for each aquifer.

2. Methodology

2.1. Model development

The AquiMod code has a modular design implemented using the object-oriented C++ programming language. It consists of three primary modules that represent the downward flux of water through the soil (root) zone and unsaturated zone and the lateral flow and subsequent discharge of groundwater through the saturated zone (Figure 1). Each module may accommodate different components based on different conceptual representations of the process being considered. Similarly, each module may be switched off entirely if the user only wishes to consider part of the hydrological system. All components for a given module adhere to the same generalised structure and exchange the same hydrological variables. Firstly, rainfall and potential evapotranspiration (PET) time-series data are fed into the soil zone module, which calculates the proportion of water that infiltrates the soil column, becomes runoff, and evapotranspires. A proportion of soil water is allowed to drain downwards to the unsaturated zone module, which attenuates the flux of recharge to a saturated zone. Finally, groundwater storage, level and discharge are calculated by the saturated zone module. It is important to note that the lumped structure of AquiMod means that the groundwater level simulations are for a single point at the observation borehole and are therefore not necessarily indicative of the levels in the system as a whole.

2.1.1. Soil zone

A single soil zone component has been selected for this study which is used as part of the Environment Agency of England and Wales' national Continuous Estimation of River Flows

(CERF) model (Griffiths et al., 2006) and been shown to simulate soil moisture fluctuations in temperate climates efficiently and comparably to more sophisticated physically-based land surface models (Sorensen et al., 2014). The method is based on the widely applied soil water balance method developed by the UN Food and Agricultural Organisation (Allen et al., 1998). This component simulates soil moisture as a function of vegetation and soil properties. The soil column is conceptualised as a bucket with a maximum volume of water available to plants after the soil has drained to its field capacity. This is termed the *total available water* (TAW) which is calculated as:

$$TAW = Z_r(FC - WP) \quad (1)$$

where Z_r is the estimated maximum root depth [L] of the overlying vegetation, and FC and WP are the soil field capacity and wilting point respectively.

As the soil moisture content decreases, it becomes more difficult for vegetation to extract moisture from the soil matrix. The proportion of TAW that can easily be extracted before this point is reached is conceptualised as *readily available water* (RAW) which is calculated as:

$$RAW = p \cdot TAW \quad (2)$$

where p is the depletion factor of the overlying vegetation.

The water balance of the soil zone is a function of the rainfall input and evaporative flux from the soil and can be written as:

$$\Delta SMD = AET - PPTN \quad (3)$$

where SMD is the soil moisture deficit [L], PPTN is the total precipitation input [L] and AET is the actual evapotranspiration rate [L] which is calculated as a function of the soil moisture deficit at the previous time-step, SMD* using the power law developed by Griffiths et al. (2006):

$$\begin{aligned} AET &= PET\left[\frac{TAW - SMD^*}{TAW - RAW}\right]^{0.2} && SMD^* > RAW \\ AET &= PET && SMD^* \leq RAW \\ AET &= 0 && SMD^* \geq TAW \end{aligned} \quad (4)$$

When the soil zone becomes saturated, a proportion of the excess water (EXW) drains to the unsaturated zone:

$$SD = BFI \cdot EXW \quad (5)$$

where SD is the soil drainage and BFI is the baseflow index. The BFI is typically used to classify low flow characteristics of river catchments (Marsh and Hannaford, 2008) and it defines the average proportion of stream flow that a river receives from groundwater discharge. It is important to note that this soil zone representation is most suited to temperate climates and may not be suitable elsewhere. However, other soil zone representations could be employed in the AquiMod software such as that devised by Rushton et al. (2006) who developed an improved version of the method used here which was shown to work well in a semi-arid setting in Nigeria.

2.1.2. Unsaturated zone

The rate at which water flows through the unsaturated zone depends on a number of factors including the permeability of the porous material, the presence of preferential pathways and the depth to the water table. To approximate the role of the unsaturated zone in attenuating the transfer of soil drainage to the water table a simple transfer function has been implemented. This method 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. In AquiMod, recharge is distributed over a number of time-steps, n, and the proportion of soil drainage for each time-step is calculated using a two-parameter Weibull probability density function:

$$f(x;\lambda,k) = \begin{cases} \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k} & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (6)$$

where k > 0 is the shape parameter and λ > 0 is the scale parameter of the distribution. The λ parameter primarily controls the location of the peak in the probability density function while k controls the density of the function around the peak (Figure 2). The resulting distribution is scaled such that the discrete integral of f is equal to unity and consequently the recharge for each time-step (R_t) is spread over the selected number of time-steps, n:

$$R_t = \alpha \sum_{x=1}^{n} f(x) \cdot SD_{t-x+1} \tag{7}$$

where α is the scaling parameter. 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. In conjunction, this method requires only three, rather than the n+1 model parameters of the Calver (1997) approach.

2.1.3. Saturated zone

The aquifer in the saturated zone module is represented as a rectangular block with dimensions Δx and Δy denoting its length and width [L] respectively. A mass balance calculation is performed at each time-step to calculate the new groundwater head:

$$R\Delta x\Delta y - Q = S\Delta x\Delta y \frac{dh}{dt} \tag{8}$$

where R is recharge input [LT^{-1}], Q is the total groundwater discharge [L^3T^{-1}], S is the storage coefficient (dimensionless), dh is the change in groundwater head [L] over time, dt [T].

This rectangular block of aquifer can be split into a number of layers of different thickness and permeability. Each layer is independent and has its own discharge outlet at the base. The total groundwater discharge is the sum of discharge from all layers in the saturated zone, which is calculated using a quadratic equation of the form:

$$Q = \sum_{i=1}^{m} T_i \Delta y \frac{\Delta h_i}{0.5\Delta x} \tag{9}$$

where i is the layer number for m layers and Δh_i [L] is the difference between the groundwater head and the elevation of the layer outlet. Importantly due to the explicit form of Equation 9 used in AquiMod, the groundwater head at the previous time-step, h^* [L] is used:

$$\Delta h_i = \begin{cases} h^* - z_i & h^* > z_i \\ 0 & h^* \leq z_i \end{cases} \tag{10}$$

where z_i is the outlet elevation. Transmissivity, T_i [L^2T^{-1}] is a function of the hydraulic conductivity, K_i [LT^{-1}] and is calculated using the following piece-wise function:

$$T_i = \begin{cases} 0 & h^* \leq z_i & i = (1 \dots m) \\ K_i(h^* - z_i) & z_i < h^* < z_{i+1} & i < m \\ K_i(z_{i+1} - z_i) & h^* \geq z_{i+1} & i < m \\ K_m(h^* - z_m) & h^* > z_m & i = m \end{cases} \tag{11}$$

When equation 9 is substituted into equation 8, the Δy term is lost. Accordingly, the saturated zone component is a lateral flow aquifer model that receives recharge from the unsaturated zone over a specified representative aquifer length (Δx). Here, Δx can be considered as the distance between a point in the aquifer where AquiMod simulates groundwater levels (typically where a field observation borehole exists) and the groundwater discharge point such as a river or a spring. By defining the saturated zone in this way, and to satisfy equation 8 it is assumed that groundwater abstractions and lateral inflows have negligible control on groundwater levels at the observation borehole. It is also important to note that the groundwater level simulations from AquiMod are for a single point at the observation borehole and are therefore not necessarily indicative of the levels in the system as a whole.

Equation 11 indicates that the chosen number and positioning of outlets in the saturated component could drastically change its behaviour. Shallow layers positioned within the zone of groundwater level fluctuation are likely to activate and deactivate as the water level rises above and falls below their outlet elevations, while layers that remain fully saturated will exhibit a more linear discharge response to changes in groundwater storage. The optimal configuration is likely to reflect the local complexities of the chosen study area such as vertical heterogeneity in the aquifer and regional discharge features such as rivers and springs. Accordingly four different saturated zone structures are presented in this study which contain varying degrees of complexity (Figure 1). The first is a three-layer representation where the outlet of the deepest layer is positioned at the base of the aquifer below the zone of water table fluctuation and can be considered to represent groundwater which flows out of the model domain via perennial flow paths. The two upper outlets are positioned within the zone of water table fluctuation and are lumped representations of surface discharge points which flow intermittently. The additionally tested saturated zone components were gradual simplifications of this model structure starting with a two-layer representation with one perennial outlet and only one intermittent outlet in the zone of groundwater level fluctuation. A simpler one-layer component was also used with a single perennial outlet at the base of the aquifer. Finally, a one-layer aquifer with a fixed transmissivity was used which represents the simplest saturated zone structure applied.

2.2. Model construction and implementation

AquiMod is an executable file that is run through the command prompt and is compatible with windows and linux machines. The driving data, parameters and preferences (such as which module components the user wishes to use) are specified using a series of text files which AquiMod reads before running. Parameter values for model structures are generated using its in-built Monte Carlo function. The run-time of AquiMod varies depending on available computer power, number of Monte Carlo runs, length of the simulation sequence and complexity of the chosen model components. For this study a PC running Windows 7 with a quad-core 2.9GHz processor was used. A run-time of 60 seconds was typically required for 10^6 runs of a 300 time-step sequence using the soil and unsaturated zone components outlined above with a three-layer saturated zone representation.

AquiMod returns the time-series of state variables or fluxes for each component including soil drainage, recharge and groundwater levels so that the user can analyse the model behaviour visually. AquiMod does not have a graphical user interface, but the output text files can easily be imported and analysed using standard external software packages. All of the outputs from AquiMod presented in this study were analysed and visualised using MATLAB.

In total five different model structure configurations were tested including the four different saturated zone components outlined previously. A fifth test was also conducted using the most efficient model structure from the prior tests, but with the unsaturated component switched off, allowing soil drainage to reach the saturated zone instantaneously. From this, the role and importance of the unsaturated zone component was also investigated.

2.3. Model Calibration

The in-built Monte Carlo functionality was used to calibrate the different model structures, where model parameters are randomly sampled from a finite parameter space to produce multiple parameter sets. One million parameter sets were sampled from a uniform distribution with upper and lower bounds defined based on expert judgment. All of the parameters used in the study are summarised in Table 1. Calibration of all of the AquiMod parameters simultaneously necessitates an infeasible number of model runs to sample the parameter space adequately, especially for the most complex model structures. Accordingly, eight of the parameters were fixed within this study using available information about the study sites. The representative aquifer length, Δx, was quantified as the distance between the observation borehole and a single discharge point on a river based on the catchment geometry and hydrogeology as an approximation of the distance to drainage. Marsh and Hannaford (2008) detail catchment BFI estimates and Boorman et al. (1995) provide distributed field capacity and wilting point values for UK soils. By analysing cross correlations between rainfall and groundwater levels, the unsaturated zone component n parameter (Equation 7) was set to the period over which there is a significant correlation at a 95% confidence level. For the five model structures, the deep outlet was set to the known

bottom elevation of the aquifer. Preliminary model runs showed that for the two and three-layer saturated zone components the remaining outlet elevation parameters significantly interacted with the hydraulic conductivity parameters. As such, a preliminary set of calibration runs were undertaken to determine elevation values that produced behavioural models to which they were subsequently set. For each model, calibration runs were performed over half of the available groundwater level time-series with the remaining half reserved to evaluate the model outside of its calibration range.

2.4. Model evaluation

Here, we have used two quantitative model performance metrics. The first is the Nash-Sutcliffe Efficiency (NSE) score (Nash and Sutcliffe, 1970), a metric which has been widely adopted by the environmental modelling community (Bennett et al., 2013) which indicates how well the model explains the variance in the observations compared with using the mean of the observations as the prediction for every time-step:

$$NSE = 1 - \frac{\sum_{t=1}^{n}(h_o^t - h_m^t)^2}{\sum_{t=1}^{n}\left(h_o^t - \overline{h_o}\right)^2} \quad (12)$$

where h_o^t and h_m^t are the observed and modelled groundwater heads at time t. A score of one denotes a perfect match to the observed data, a value of zero indicates that the model is as efficient as using the mean and a negative score is less efficient than this. This NSE was used to compare the relative efficiency of the different model structures and parameter sets in order to determine the optimum model configuration. To complement the NSE, monthly and overall bias metrics were also calculated and plotted to identify systematic deficiencies in model predictions which lead to over or underestimation of groundwater levels.

It should be noted that the term 'optimum' is used here with the knowledge that there could be other models which are equally, or even superior predicting tools (Beven and Freer, 2001). It should also be noted that simple empirical indicators of model fit such as the NSE have come under criticism in the past for returning high efficiency scores even when

simulations show significant magnitude and timing errors (Legates and McCabe, 1999; Pappenberger et al., 2004). Furthermore, Beran (1999) revealed that the NSE possesses in-built biases that can exaggerate model skill. Accordingly, a qualitative comparison of the simulated and observed hydrographs has also been conducted, as recommend in the review of characterising model performance by Bennett et al. (2013) to complement the NSE and bias indicators to provide further insight into the relative strengths and weaknesses of AquiMod.

2.5. Study Sites and Data

To assess the performance of the model, AquiMod has been applied to four aquifers with observation boreholes in the UK within the different lithologies of the Cretaceous Chalk, Carboniferous Limestone, Cretaceous Lower Greensand, and Triassic Sandstone. The climate of the sites are typical of the UK with wet and cold autumn and winter months between October and March and drier, warmer spring and summer months between April and September. The four study sites have been chosen because of their contrasting hydrogeological settings, but also because the observation boreholes are located away from significant groundwater abstractions. Available groundwater level time-series for each borehole have been obtained from the National Groundwater Level Archive (Marsh and Hannaford, 2008) between 1961 and 2005 and are reported as meters above sea level (m asl). Groundwater level monitoring is typically undertaken on a weekly or monthly basis, but the frequency is variable. Due to the irregularity of measurement these data have been converted to monthly time-series using linear interpolation (Figure 4). Accordingly, AquiMod has been configured to run on a monthly time-step.

While no pumping tests have been carried out at these observation boreholes specifically, some information on regional transmissivity and specific yield values does exist; those reported below are taken from Allen et al. (1997).

The Chilgrove House observation borehole is located in the River Lavant catchment in south-east England (Figure 3). The hydrograph has an annual sinusoidal appearance, although double and higher multiple peaks are relatively frequent, generally due to the uneven temporal distribution of rainfall. Flow within the saturated zone of the chalk occurs

predominantly in the upper 50 m of the profile through primary and secondary fractures. Hydraulic conductivity is generally highest in the zone of water table fluctuation and in valleys where fractures have been developed by dissolution. The median transmissivity for the chalk in this area is 440 $m^2 d^{-1}$, and the 25^{th} and 75^{th} percentile values are 230 and 1600 $m^2 d^{-1}$ respectively. Specific yield values of the chalk are typically in the range 0.5-2%. The BFI of the River Lavant at Graylingwell, 9 km south-east of the borehole, is 0.82 (Marsh and Hannaford, 2008). The mean depth to groundwater level is 28.6 m. It is generally accepted that fluxes within the unsaturated zone are transmitted through the matrix until they exceed the saturated hydraulic conductivity of the matrix, at which point fracture flow becomes dominant (Ireson et al., 2006). Previous studies have shown, however, that the generation of fracture flow is rare and for the majority of the time fluxes are transmitted by the matrix (Mathias et al., 2006).

The Hucklow South observation borehole is in the River Wye catchment, which drains the Carboniferous Limestone in central England. Geological logs of this 123.6 m deep borehole do not exist but it is considered that it is likely to have been drilled down to the Litton Tuff and Cressbrook Dale Lava members of the Peak Limestone Group, which form an effective base to the unconfined aquifer (Downing et al., 1970). Groundwater levels have fluctuated by approximately 30 m over the period for which observations have been obtained from 1969 to 2005. The minimum groundwater level over this period is approximately 55 m below ground level, but the rest water level is 16 m below ground level on average. Two or more peaks are frequently observed in the winter months due to rapid response to rainfall. The BFI of the River Wye at Ashford, 8 km south of the borehole, is 0.75 (Marsh and Hannaford, 2008). There is limited information on hydrogeological properties of the Peak Limestone Group, but from the few tests that have been conducted, specific yields typically range from 0.5 to 8%. Only six pumping tests have been conducted in this region, yielding transmissivity values ranging from 0.1 to 770 $m^2 d^{-1}$. All but one of these tests yielded transmissivity values less than 60 $m^2 d^{-1}$.

The Lower Barn Cottage observation borehole is located in the River Ouse catchment in south-east England. The borehole has been used to monitor groundwater levels in the

unconfined Lower Greensand, an important aquifer in south-east England, since 1975. At Lower Barn Cottage the Greensand formation is shallow with a total depth of approximately 9 m, and the annual mean groundwater level resides only 2.6 m below the surface. The Lower Greensand comprises a complex series of variably cemented clays and sands, the heterogeneity of which results in a relatively irregular hydrograph. The median value of Lower Greensand transmissivity estimates is 270 $m^2 d^{-1}$, and the 25^{th} and 75^{th} percentile values are 140 and 500 $m^2 d^{-1}$ respectively. Specific yield values are typically in the range 10-20%. The Lower Greensand outcrops over a small proportion of the Ouse catchment and so a locally-related BFI cannot be identified from a nearby gauging station for the aquifer in this region. A generally representative, BFI of 0.8 has been estimated for the Lower Greensand by Bloomfield et al. (2011) using data from the UK Hydrometric Register (Marsh and Hannaford, 2008).

The Skirwith observation borehole is used to monitor groundwater levels in the St Bees Sandstone formation located in the River Eden catchment in north-west England. The aquifer is largely unconfined although the Skirwith borehole log indicates some localised confinement by glacial boulder clay deposits. Regular groundwater level measurements have been taken since November 1978 although no records between December 2000 and March 2002 exist due to site access restrictions as a result of the foot and mouth disease outbreak in the United Kingdom. Two river flow gauging stations are located within 4 km of the Skirwith observation borehole on the River Eden at Udford and Temple Sowerby with an average BFI of 0.43 (Marsh and Hannaford, 2008). Only seven pumping tests have been conducted in this formation which yielded transmissivity values from tens up to 2000 $m^2 d^{-1}$ with a geometric mean of 100 $m^2 d^{-1}$. No information on typical specific yield properties in this area exists.

Monthly PET time-series have been extracted from the Meteorological Office Rainfall and Evaporation Calculation System (MORECS) (Field, 1983). This calculates PET on 40×40 km grid from synoptic station data using a modified version of the Penman-Monteith equation (Monteith and Unsworth, 2008). Rainfall data have been extracted from a 1×1 km gridded

dataset (Keller et al., 2006) constructed for the Environment Agency of England and Wales' CERF model (Griffiths et al., 2006) using data from the UK network of rain gauges.

3. Results

First we present an analysis of the most efficient model structures calibrated for each study site by comparing the simulated and observed groundwater level time-series. Next, we describe the sensitivity and identifiability of each calibration parameter for these most efficient structures. Finally, we show the effect of using progressively simpler representations of saturated and unsaturated groundwater flow using the five different model structures outlined.

3.1. Groundwater level simulation efficiency

First we analysed which of the model structures could simulate the groundwater level time-series most efficiently. This gave a general overview of the relative strengths and weaknesses of AquiMod for simulating groundwater level hydrographs in the four contrasting study catchments, and it also provided a benchmark against which the remaining model structures could be compared. Table 2 lists the NSE scores obtained for the calibration and evaluation runs for each model structure. The NSE and bias results are also included for the entire simulation sequence (combined calibration and evaluation periods).

For the Chalk site, the three-layer aquifer representation returned the highest combined NSE of 0.91, although the bias was 0.16 m greater than in the two-layer model. However, this difference in bias constitutes <0.4% of the range of observed levels recorded at this borehole and therefore the three-layer representation was deemed optimal. The high efficiency score is reflected in the hydrograph where the model closely matches the levels in the observation record throughout the simulation period (Figure 5a). There are some localised sections of the hydrograph when the model is less efficient including the winters of 1974/1975 and 1997/1998 when it underestimates the peak levels by up to 7 m. Even so, a comparison of the simulated and observed mean monthly groundwater levels (Figure 5b) shows that on average this model structure can capture the seasonality of the hydrograph

extremely well both in terms of timing and magnitude. The monthly biases are small with respect to the variability of the hydrograph, with the largest bias of 1.2 m shown in December and an overall positive annual bias of 0.25 m.

The three-layer groundwater component was also the most efficient structure for the Limestone aquifer for which it produced the smallest bias. It was considerably less efficient than the Chalk model with a combined NSE of 0.65, and the NSE decreased significantly (-0.09) between the calibration and evaluation runs. In particular, the model is unable to capture the flashy response to recharge when groundwater levels are high, sometimes adding extra features or missing them entirely. For example, in the winter of 1986/1987 there is only one observed peak, but the model simulates two separate ones (Figure 5c). Conversely, during the winter of 1990/1991 there are two observed peaks, but the model only captures one of them. However, a comparison of observed and simulated mean monthly groundwater levels (Figure 5d) shows that AquiMod is able to capture the groundwater level seasonality with reasonable accuracy. The largest bias is seen over the summer months when the model overestimates the low groundwater levels in July by 0.9 m on average, a defect that is clearly seen in the simulated time-series.

For the Lower Greensand aquifer, the two-layer groundwater component returned a calibration NSE of 0.73 and a negligible bias. The NSE increased to 0.93 over the evaluation sequence, although it should be noted that some of the largest errors are also observed over this period. The greatest discrepancy between the observations and simulations occurred during the winter of 2000/2001 when the model underestimated the groundwater level peak by 0.58 m (Figure 5e). Even so, AquiMod is able to capture some of the longer inter-annual signals that are present over the evaluation period such as the prolonged low levels between 1990 and 1993 and between 1996 and 1997. It is also able to replicate the average timing of the hydrograph reasonably well, capturing the September minima, although predicting the mean groundwater level peak a month later than observed (Figure 5f). The largest monthly biases are seen during the recharge season between September and December when it underestimates average monthly groundwater levels by as much as -0.09 m.

Finally, for the Sandstone aquifer, the two-layer saturated zone component was the most efficient at replicating the hydrograph, achieving a combined NSE of 0.85 and the joint smallest bias of -0.05 m. As with the calibrated model of the Lower Greensand aquifer, the evaluation run NSE is higher than that for the calibration run even though the largest errors are seen over the evaluation sequence. Here, the model underestimates multiple peak winter levels (Figure 5g), with the largest discrepancy during the 1994/1995 winter (-0.6 m). Nevertheless, AquiMod is able to capture the seasonality and timing of the hydrograph well (Figure 5h), reproducing the mean monthly peak groundwater level in March and the lowest mean level in October with relatively small biases in comparison to the average variability of the hydrograph.

3.2. Sensitivity Analysis

To extend the assessment of the most efficient model structures for each site, a sensitivity analysis of the model parameters was conducted to determine which parameters influence the model efficiency, which parameters are identifiable and if these can be related to known physical characteristics of the study sites, thereby allowing them to be constrained based on available catchment information. A series of dotty plots have been constructed from the output files produced by AquiMod (Figure 6), which show how the model efficiency changes as each parameter is perturbed.

For the soil zone, the Zr parameter has the most influence on model efficiency, although the sensitivity of this parameter differs for each study site. The models of the Chalk and Limestone aquifers both demonstrate a sloped response surface, indicating that Zr is identifiable. In contrast, the response surface for the Lower Greensand model, with two separate peaks, and the Sandstone model, which flattens off, do not show a clear unique optimum. The depletion factor parameter, p which controls the point at which the evaporation rate falls below the potential rate is relatively insensitive for all sites. For the models of the Chalk, Limestone and Lower Greensand, there is a gradual upward trend in the response surface as p decreases which indicates some sensitivity to this parameter, while for the model of the Sandstone, p is insensitive.

Sensitivity in the unsaturated zone arises almost exclusively from the λ scale parameter which controls the location of the peak in the Weibull distribution and which represents the time taken for soil drainage to pass through the unsaturated zone module. For the Chalk, Limestone and Lower Greensand models, the simulations are most efficient when λ is less than 1.5. These values produce a Weibull distribution with a peak close to one, causing the majority of soil drainage to pass through the unsaturated zone during the first time-step (instantaneously). Figure 7 shows that for these models between 73% (Limestone) and 100% (Lower Greensand) of soil drainage is allowed to pass through the unsaturated zone instantaneously. For the Sandstone aquifer there is a clear maximum in the λ response surface at a value of 2.1. Here, soil drainage is attenuated more, so that the majority (47%) recharges the groundwater a month later. The model of the Lower Greensand is the only one that shows any sensitivity to k, which controls the concentration of recharge around the peak, where higher values (which result in very concentrated recharge over a single time-step) are optimal. The remaining models are not sensitive to the k parameter.

The calibrated unsaturated zone components indicate that the Sandstone unsaturated zone has the greatest lagging affect on groundwater recharge. For comparison, a cross correlation analysis between the rainfall and groundwater level time-series was performed to estimate the peak response time of groundwater to rainfall at each site. That is, the time taken for the majority of an instantaneous flux of rainfall to reach the water table and perturb the groundwater storage. It should be noted that the groundwater level time-series were de-seasonalised using the Loess method (Cleveland et al., 1990) to remove the signal induced by seasonal evaporation fluxes. For the Chalk, Limestone and Lower Greensand study sites, the highest cross-correlation coefficients were obtained for lead-lags between zero and one month while for the Sandstone site the highest cross-correlation score was obtained for a higher lead-lag of two months.

The specific yield parameter, S, is identifiable for all of the study sites. Optimum values of 0.6%, 0.7% and 22.3% were obtained for the Chalk, Limestone and Lower Greensand, respectively, all which conform to values obtained from field pumping tests (Allen et al.,

1997). For the Sandstone model, the optimum S value of 8.3% was obtained although there are no data available for this aquifer against which to compare this value. The conductivity parameters all demonstrate sensitivity and all are identifiable. Using these parameters and the maximum, minimum and mean simulated groundwater levels over the calibration and evaluation periods, the corresponding maximum, minimum and mean model transmissivity values were calculated for each study site for comparison to available aquifer data. For the model of the Chalk, the transmissivity ranges between 24 – 720 $m^2 d^{-1}$ with a mean of 178 $m^2 d^{-1}$. The modelled transmissivity for the Limestone aquifer is less variable, ranging between 17 – 418 $m^2 d^{-1}$ with a mean of 148 $m^2 d^{-1}$, while the model of the Sandstone transmissivity values range between 1 – 202 $m^2 d^{-1}$ only with a mean of 94 $m^2 d^{-1}$. The Lower Greensand site has the lowest transmissivity values, ranging from 7 – 119 $m^2 d^{-1}$ with a mean of 39 $m^2 d^{-1}$. All of these conform to those values obtained from pumping test data, although it should be noted that for the Lower Greensand model, the transmissivity values fall within the lower quartile of available data. The model of the Limestone is the only one where the calibrated K values do not increase with elevation. Instead, the intermediate layer returns the highest conductivity, which may indicate a localised zone of high transmissivity between the middle (254.9 m asl) and upper (262.6 m asl) outlets.

3.3. Model Structure Analysis

In addition to evaluating the most efficient structures obtained for each site, a final analysis was conducted to explore the impact of using progressively simpler representations of the saturated and unsaturated zones on the simulated results and to provide further insight into the controls on groundwater storage and discharge at each site.

First, the calibration of the models using the four different saturated zone components has been compared. Figure 8 shows the simulated hydrographs over the evaluation sequence using these four different structures. It indicates that the difference between using the two and three-layer aquifer representations for the Chalk, Limestone and Lower Greensand study sites are subtle. For the Chalk site, both the two and three-layer calibrated models return very similar combined efficiency scores of 0.90 and 0.91 respectively and almost identical simulations (Figure 8a). There are slightly more pronounced differences for the

Lower Greensand site where the three-layer groundwater component underestimates levels more so than the two-layer representation between 1991 and 1994, and in 2001 (Figure 8c). For the Limestone catchment the three-layer component returns only a marginally higher (+0.03) combined NSE than the two-layer model. However, the simulation bias improves by an order of magnitude (Table 2). The reason for this can be seen by comparing the two simulation hydrographs (Figure 8c) where the two-layer aquifer representation overestimates the peak groundwater levels substantially. This suggests that the inclusion of the intermediate high conductivity layer benefits the simulation efficiency by removing this positive bias. For the Sandstone site, the two and three-layer models achieve similar calibration NSE scores, but the two-layer representation returns a much higher NSE over the evaluation period (+0.34). Indeed, the efficiency of the three-layer component deteriorates between 1996 and 2000 where it underestimates the groundwater levels while the simpler two-layer groundwater component captures the behaviour of this hydrograph well even over this period.

Simplifying the groundwater component further to a one-layer aquifer representation resulted in a significant loss of simulation efficiency for all study sites. The one-layer variable transmissivity representation for the Chalk still returned a high calibration NSE of 0.73 although this fell significantly to 0.49 for the evaluation run indicating that this structure does not adequately characterise this groundwater system. Certainly, while the model is still able to capture the seasonality of the hydrograph, the suitability of this structure breaks down during periods of exceptionally low levels (e.g. 1973) and high levels (e.g. 2001) where the model exaggerates the magnitude of these events (Figure 8b). By simplifying the groundwater component further to a one-layer representation with a constant transmissivity, these deficiencies are amplified. Similar behaviours for both one-layer representations are observed for the Lower Greensand site (Figure 8f), especially between 1992 – 1995 where underestimation of levels persists and between 2001 – 2005 where an overestimation of levels persists. In this case, the lack of a high conductivity layer in the zone of fluctuation causes the model to deviate from the observed levels for extended periods of time. This behaviour occurs most dramatically in the Sandstone simulations after 1997 (Figure 8h). For the Limestone site, using a one-layer model has a completely different

affect on the simulations. Here, it appears the model is not able to capture any of the variability in the hydrograph. Actually, the calibration NSE scores which are close to zero for both the one-layer variable and fixed-transmissivity groundwater components indicate that the best the model can do is approximately match the mean of the observations.

The fifth calibrated model structure used the optimum groundwater component from the previous tests but with the unsaturated zone component removed all together. This is likely to impact the timing and seasonality of the simulated hydrographs as the unsaturated zone component attenuates and translates soil drainage to the water table. To investigate this, the simulated annual groundwater level distributions using model structures with and without the unsaturated zone component were compared (Figure 9). Here, it can be seen that the average timing of the Chalk, Limestone and Lower Greensand hydrographs do not change considerably. However it should be noted that for the Limestone study site, removing the unsaturated zone allows the model to correctly simulate the average peak groundwater levels in January rather than December. Of course, the role of the unsaturated zone module for each of these sites has shown to be minimal (Figure 7), but even so, removing it does have some impact on the monthly errors of these models especially for the Limestone and Lower Greensand models which show an enhanced positive and negative bias respectively. The removal of the unsaturated zone component results in a significant drop in overall efficiency from 0.85 to 0.76 for the Sandstone model. Here, the timing of the peak and trough of the hydrograph is out by one month (Figure 9d) and there is a consistent overestimation of levels between October and January, and an underestimation of levels between February and August.

4. Discussion

4.1. Groundwater level simulation efficiency

AquiMod can simulate groundwater level time-series in contrasting aquifers with considerable accuracy. For the quasi-sinusoidal Chilgrove House Chalk hydrograph, AquiMod was able to capture the seasonality, timing and magnitude of the peaks and troughs very efficiently, achieving the overall highest NSE score. Similarly, for the more slowly responding, but still largely sinusoidal Skirwith Sandstone hydrograph, AquiMod was very

efficient over the calibration and evaluation sequences. The most efficient calibrated model for the Lower Barn Cottage observation borehole in the Lower Greensand aquifer returned the highest NSE of any model over the evaluation sequence. AquiMod was able to closely match the variability in this observation record, both in terms of the intra-annual seasonality and the longer term fluctuations, an observation which is encouraging given the irregularity of the hydrograph resulting from the heterogeneous nature of the Lower Greensand formation. AquiMod was less efficient at capturing the complex behaviour of the hydrograph in the Limestone aquifer achieving the lowest combined NSE of all the study sites. Here, the groundwater level, storage and discharge behaviour appears to be more complex than the other sites, where the hydrograph shows a smooth recession, but rapidly fluctuates during the winter months when levels are high. This behaviour is indicative of the complex flow pathways caused by the Limestone's low primary porosity which means that flows are restricted to fast preferential fracture pathways that are only activated at certain times of the year when water saturates them (Atkinson, 1977). Even using the most efficient three-layer saturated zone representation, AquiMod was not able to capture this complex behaviour and showed a notable strong positive bias during the summer minimum levels.

Of course, deficiencies in simulations, like those observed for the Limestone model, may arise from a number of sources including errors in the meteorological data used to drive the models and the groundwater level observations used to evaluate them. The impact of running the simulations on a monthly time-step and the resultant smoothing of the driving rainfall and PET data should also be considered. By doing so, the short, intense rainfall events that can lead to significant recharge fluxes are not captured by the model which may result in a significant underestimation of recharge (Howard and Lloyd, 1979). For the Limestone model, the relatively low efficiency score compared to the other study sites suggests there are also shortcomings in the model structure that means it is not able to capture the non-linear storage-discharge relationship of this aquifer adequately. It is certainly possible to conceive new model structures that may improve upon the simulations described in the study. Certainly, the object-oriented structure of the AquiMod code allows for new structure representations to be included easily. One feature that has not been included in the saturated zone components used here is a variation in storage coefficient

with depth, which has been shown to be important to include in some groundwater modelling studies of the Chalk and Jurassic Limestone aquifers in the UK (Rushton and Rathod, 1981; Rushton et al., 1982).

Structural inadequacies aside, another issue made apparent during the initial efficiency evaluation of AquiMod was its inability to capture new events outside of those observed in the calibration sequence. Even for the most efficient model of the Chalk aquifer, it was not able to simulate extreme wet events in the evaluation sequence with a tendency to underestimate the response to them. A subsequent analysis of the rainfall dataset showed that these periods were especially wet, and wetter than any of the winters contained in the calibration dataset. The issue of model robustness over contrasting climatic conditions is important, as these can lead to significant simulation uncertainties outside of the chosen calibration period. It is acknowledged that the adopted approach of selecting a single calibration and evaluation time-series does not account for these uncertainties, and while beyond the scope of this paper, more rigorous approaches to quantify these uncertainties do exist such as the generalized split-sample test (GSST) formulated by Coron et al. (2012) which would be relatively straight forward to use in conjunction with the AquiMod software. It is also acknowledged, that the choice of objective function used to evaluate the model efficiency can greatly influence the assessment of the model and subsequently the selection of suitable structures and parameter sets. For this study, the NSE was chosen to assess the simulation efficiency, which was deemed adequate for this initial demonstration of the AquiMod software. However, as discussed previously, the NSE has its shortcomings (Beran, 1999; Legates and McCabe, 1999; Pappenberger et al., 2004). Certainly, the increase in simulation efficiency over the evaluation period for the Lower Greensand and Sandstone models is probably more reflective of the increased variability in the observed hydrograph rather than a reduction in residuals which were actually larger over the evaluation sequences. Accordingly, the choice of objective function should reflect the purpose of the modelling exercise (Bennett et al., 2013).

4.2. Identification of suitable model parameters and structures

The modular structure of the AquiMod software, in addition to its in-built Monte Carlo parameter sampling functionality, has allowed the sensitivity of the groundwater level simulations to be assessed in response to changes in the calibration parameters and model structure. Each of the four modelled sites have shown very different behaviours in response to these perturbations. Here, we discuss the results from these analyses by considering each of AquiMod's three modules in turn and their influence on the behaviour of the models of each study site.

The soil zone module provides an important role in AquiMod by partitioning rainfall into runoff, evapotranspiration, soil storage and drainage. Even so, the depletion factor (p) parameter was remarkably insensitive while the maximum root depth (Zr) parameter was only identifiable for the Chalk and Limestone models. This finding is surprising given that these parameters control the evaporation rate from the soil, and thus can dramatically alter the recharge input to saturated zone module. However, it is conceivable that other model parameters could interact with p and Zr. For example, if they are configured so that the overall recharge flux is reduced, this could be compensated for by reducing S in the saturated zone module to maintain the variability in the hydrograph. A subsequent set of runs were conducted for the four study sites which showed that if all other parameters are fixed, the soil parameters do have a significant impact on groundwater level simulations, indicating that parameter interaction can explain their insensitive behaviour for this study. These findings suggest that it may in fact be beneficial to fix these using the best available soil and vegetation information. Certainly, available large-scale soil datasets such as the Land Information System (LandIS) for England and Wales (Proctor et al., 1998) and the Harmonised World Soil Database (FAO et al., 2012) can be used to inform the selection of these parameters.

The Weibull unsaturated zone component influences the behaviour of AquiMod greatly. It is calibrated using two calibration parameters: λ which approximately controls the lag in the unsaturated zone; and k which controls the attenuation (spread) of the recharge response to soil drainage. Both showed varying degrees of sensitivity and optimal values across the

study sites. Of the two, λ was most sensitive as it controls the seasonal signal in the hydrograph simulation and therefore has a significant impact on the overall efficiency of the model. The Weibull distributions for the Limestone and Chalk models both resulted in a rapid flux of soil drainage to the water table, although a small portion (approximately 20%) of soil drainage was lagged by one month. However, the k parameter was found to be insensitive for these models, suggesting this spread of recharge after the peak is not uniquely optimal. Rather the rapid peak response induced by the optimal small λ values was most crucial to the efficiency of these models. For the Limestone aquifer, the reason for this rapid response is likely to be percolation through preferential fracture pathways. For the Chalk this rapid response to rainfall is not uncommon even where the unsaturated zone has a significant depth (28.6 m on average for the Chilgrove House borehole). This is because of so-called 'piston flow' through the Chalk matrix in the unsaturated zone. Here, groundwater recharge occurs as a result of a pressure wave traversing through the, generally highly saturated, matrix which displaces water from bottom of the unsaturated zone rather than percolating through it (Mathias et al., 2005).

Soil drainage passed instantaneously to the underlying saturated zone module in the most efficient model of the Lower Greensand aquifer. The sensitivity analysis showed that the most efficient simulations were produced when λ was small (< 1.5) and k was large (> 4), resulting in a rapid and sharp recharge response. In fact, at this site the water table lies only 2 m below ground level on average, which explains the instantaneous flux of soil drainage to the saturated zone. Incidentally, this was the only model that showed any sensitivity to k suggesting the model fit was very dependent on this sharp spiky recharge response. It should be noted that for all models, the sensitivity of k is likely to be inhibited by the coarse monthly time-stepping employed for this study as much of the definition of the Weibull distribution is lost once it is scaled over a small number of monthly time-steps (clearly seen in Figure 7). Therefore, k should show more sensitivity if smaller time-steps are used.

Due to the small attenuation affect of the optimized Weibull unsaturated zone components for the Chalk, Limestone and Lower Greensand models, the impact of removing this module all together from the AquiMod structure had little effect on the overall simulation efficiency.

However, for the Sandstone model which returned the largest optimum λ value resulting in the greatest attenuation of recharge, removing the unsaturated zone component had a significant impact on the efficiency of the model as it could no longer match the timing of the groundwater level hydrograph. The cross-correlation analysis between rainfall and de-seasonalised groundwater levels indicated a longer response time for the Sandstone aquifer and consequently, the importance of including the unsaturated zone module for this site. Interestingly, the average thickness of the unsaturated zone for the Sandstone site is only 3 m which presumably would have little impact on the soil drainage flux. Actually, while the Skirwith borehole groundwater catchment is known to be primarily unconfined, the drill log for this borehole revealed that there is a potentially confining clay layer present up to 3 m thick. This suggests that the lagged response to rainfall determined from the calibrated model and from the independent cross-correlation analysis could actually be a result of the time taken for groundwater levels in the borehole to respond to recharge taking place further away where the aquifer is unconfined. This implies that it is not possible to say with certainty that the lag-response induced in AquiMod by the Weibull unsaturated zone component represents the time delay between water draining from the base of the soil to the water table exclusively. Rather, this component is a transfer function that introduces a degree of memory into AquiMod, even though this may derive from a number of sources. For example, the saturated zone itself will have a certain amount of memory in its response to recharge related to the flow and storage characteristics of the aquifer in question (Kooi and Groen, 2003; Neuzil, 1986).

Interpretation of the parameters in AquiMod should always be made with care. AquiMod is a lumped model and as such the calibrated parameter sets and structures represent simplified conceptualisations of heterogeneous field conditions. Certainly, questions remain as to where the lag induced by the unsaturated zone module for the Sandstone model originates, and whether it is in fact an artefact of a distribution of localised confining layers overlying the aquifer which cannot be explicitly represented in AquiMod. Some of the parameters, such as those used in the soil zone component show little or no sensitivity, and there is evidence of parameter interaction, both of which inhibit the confidence that we can have in the uniqueness of the model parameters and their relation to field properties. It is

also important to consider the model assumptions. For example, for the saturated zone component, an aquifer length was defined which assumes a fixed discharge point and negligible lateral inflows. In highly dynamic catchments where the groundwater divide moves significantly, these assumptions are likely to break down. Even so, comparison to available field data has shown that the calibrated specific yield and hydraulic conductivity parameters of the most efficient model structures were identifiable and were consistent with available field data for all sites which implies that they have some physical relevance and that it might be possible to constrain the bounds of the parameter space before calibration using measurable catchment properties. The calibrated Lower Greensand transmissivity values fell into the lower quartile of the 40 pumping test estimates, which is likely to be an artefact of the relatively thin Lower Greensand in this region (9m in comparison to a maximum thickness of 220m) as transmissivity is the integral of permeability over depth. The optimum transmissivity obtained for the Chalk aquifer also fell within the lower quartile of the observed data for this region.

While the calibrated two and three-layer groundwater components for the Chalk, Limestone and Lower Greensand models produced models that were almost as efficient at simulating their respective groundwater level hydrographs, the three-layer component for the Limestone did reveal some interesting features. Here, the three-layer model was considerably better than the two-layer model over the evaluation sequence. Furthermore, it was the only model structure to include an intermediate high conductivity zone. This helped to reduce an overall positive bias in the model by over an order of magnitude compared to the two-layer component. Physically, the presence of a high conductivity zone is plausible given the complex fracture flow known to dominate this aquifer as previously discussed, and as such it may be that the addition of layers to the AquiMod saturated zone module may help to test concepts of preferential flow pathways.

The model of the Sandstone showed the most significant change in simulation efficiency when switching between a two and three-layer aquifer representation. Here two layers were optimal, particularly over the evaluation sequence. Theoretically, a three-layer aquifer representation should be able to simulate the storage-discharge relationship as efficiently as

a two-layer component: one only needs to understand the mathematical formulation (Equation 11) to see that this is the case. So the argument that the three-layer representation is mathematically less suited to the Sandstone aquifer than the two-layer representation is not adequate. Instead, a more likely reason for this loss in efficiency with complexity is an issue of parsimony. It appears that the three-layer component was over-fitted to the calibration sequence and subsequently performed worse than the simpler, more parsimonious, two-layer structure. In this sense, the comparison between the two highlights an important consideration when applying these simple models to different aquifers; the most complex model will not always prevail as the most suitable model structure.

The impact of simplifying the model structure further to single-layer, variable and constant-transmissivity representations, resulted in a sharp fall in simulation efficiency for all sites although the impact for each site was contrasting. The most dramatic impact on the simulations was observed for the Limestone aquifer where the optimization procedure could only produce a model that approximately matched the mean of the observations. It is worth noting that theoretically, the amplitude of the hydrograph could have been matched, but probably not the pattern of fluctuation, by extending the calibration range of the storage coefficient to much smaller values. However this could not have been justified physically. Rather, this serves to further highlight the complexity of the discharge response to changes in aquifer storage at this site. The Chalk showed the highest efficiency scores using the one-layer aquifer representations, although the suitability of these simple structures was shown to break down during periods of exceptionally high and low levels. A similar pattern was observed for the Lower Greensand site with persistent under and overestimation of levels and most dramatically for the Sandstone site. Indeed, the presence of at least one high conductivity layer in the zone of fluctuation acts as a sort of discharge moderator where it drains rapidly when the groundwater levels are high and ceases to flow when the level falls below it. This combination prevents the model from deviating from the typical range of levels for the aquifer. More importantly though, the gradual simplification of the model structure from three layers to a single-layer fixed transmissivity component is a gradual linearization of the storage-discharge relationship, a relationship that is known to be

non-linear because of vertical heterogeneity in the storage and hydraulic conductivity properties of the aquifer and other regional characteristics such as the presence of river and spring discharge points. As such, these one layer representations are unlikely to be useful for most applications of AquiMod.

The inclusion of extra complexity, be it through the addition of extra layers in the saturated zone module or other modifications can improve model efficiency and provide further insight into the controls on groundwater level fluctuations. However, the inclusion of extra parameters may also introduce more uncertainty into the model and result in an unparsimonious model structure (Jakeman and Hornberger, 1993). Interestingly, for three of the study sites, the act of removing the unsaturated zone component, thereby simplifying the structure of the model, had a small impact on the model efficiencies. It may also be beneficial to use simpler functions that describe the relationship between groundwater storage, level and discharge that require fewer parameters such as the cubic function employed by Moore and Bell (2002), or even simpler approaches such as the semi-analytical solution presented by (Park and Parker, 2008) who assume recharge is a fixed proportion of rainfall, and lump the flow, storage and hydraulic gradient variables into a single parameter assumed constant over space and time. This of course deviates from the layered saturated zone structure used in AquiMod, and most physically based models, due to the discrete step-wise layering in flow and storage properties often observed in groundwater aquifers (Cross et al., 1995; Rushton and Rao, 1988). However, where the variation of aquifer properties, such as the hydraulic conductivity, are known to vary gradually, it may also be possible to include statistical distributions of these properties which are characterised by fewer parameters.

The subject of prediction uncertainty is one that has not been addressed in this paper, although the issue has arisen on a number of occasions not least because of issues of non-uniqueness of acceptable model parameter sets and structures. Certainly, future applications of AquiMod should seek to take advantage of its fast run-time and its ability to incorporate multiple model structures to rigorously quantify uncertainty. Approaches such as the previously mentioned GSST scheme and well established multi-model procedures

such as the Generalised Likelihood Uncertainty Estimation (GLUE) method (Beven and Binley, 1992), which allows the user to quantify model parameter and structural uncertainty, are increasingly being applied in hydrological modelling applications. The work of Clark et al. (2008), who devised the Framework for Understanding Structural Errors (FUSE) in rainfall runoff models, has demonstrated the importance of understanding structural uncertainty for river flow simulations, and the AquiMod software could provide a useful platform to extend this type of analysis to groundwater level simulation applications in the future.

5. Conclusions

AquiMod is a simple, lumped conceptual groundwater level prediction tool that can be run quickly and efficiently to simulate groundwater levels for contrasting aquifer types. It has been shown to be very efficient at simulating smooth quasi-sinusoidal groundwater level hydrographs in the Chalk and Sandstone and the highly irregular hydrograph of a Lower Greensand observation borehole, but in its current form appears to be less suited to complex fractured aquifers such as the Carboniferous Limestone where it could not capture summer low levels and the rapid fluctuations in the winter high levels adequately.

The AquiMod software is simple to use, making it more accessible to hydrogeologists than more sophisticated, and often complex, distributed physically-based models. It also has the potential to derive some average aquifer flow and storage properties and could be used as a preliminary investigative tool before applying a distributed model. The software allows the user to experiment with different model structures and sample multiple parameter sets, which is especially useful for simulating groundwater levels at sites with storage-discharge relationships that cannot necessarily be characterised by a single conceptual model. Indeed, through using progressively simpler representations of aquifer properties, this work has highlighted the importance of incorporating vertical heterogeneity of aquifer hydraulic conductivity in the groundwater model structure to capture aquifer storage-discharge dynamics efficiently.

Due to its modular structure, new components can be included in AquiMod easily, which combined with its in-built Monte-Carlo parameter sampling functionality could provide a platform from which more rigorous data-based mechanistic modelling approaches (Young et al., 2007) can be implemented in the future. There is also the potential to expand its application to investigate issues of parameter and structural uncertainty. Here, similar studies for river flow simulations such as those conducted by Beven and Binley (1992), Clark et al. (2008) and Coron et al. (2012) should help to guide the focus of these future applications.

6. Acknowledgements

The code was developed using core science funds of the Natural Environment Research Council (NERC) British Geological Survey's (BGS) Groundwater Science and Environmental Modelling Directorates. The groundwater level data for this study were taken from the NERC BGS National Groundwater Level Archive. The climate data were made available by the NERC Centre for Ecology and Hydrology, and the UK Met-Office. We thank the anonymous reviewers for their very helpful comments on the paper. The authors publish with the permission of the Executive Director of the British Geological Survey.

7. References

Adams, B., Bloomfield, J. P., Gallagher, A., Jackson, C., Rutter, H. & Williams, A. 2008. FLOOD 1. Final Report. *British Geological Survey Open Report (OR/08/055).*

Aflatooni, M. & Mardaneh, M. 2011. Time series analysis of ground water table fluctuations due to temperature and rainfall change in Shiraz plain. *International Journal of Water Resources and Environmental Engineering*, 3**,** 176-188.

Ahn, H. 2000. Modeling of groundwater heads based on second-order difference time series models. *Journal of Hydrology*, 234**,** 82-94.

Allen, D. J., Brewerton, L. J., Coleby, L. M., Gibbs, B. R., Lewis, M. A., MacDonald, A. M., Wagstaff, S. J. & Williams, A. T. 1997. The physical properties of major aquifers in England and Wales. *British Geological Survey Technical Report (WD/97/34). Environment Agency R&D Publication 8.*

Allen, R. G., Pereira, L. S., Raes, D. & Smith, M. 1998. Crop evapotranspiration - Guidelines for computing crop water requirements - FAO Irrigation and drainage paper 56. Food and Agriculture Organization of the United Nations.

Alley, W. M., Healy, R. W., LaBaugh, J. W. & Reilly, T. E. 2002. Flow and Storage in Groundwater Systems. *Science*, 296**,** 1985-1990.

Anaya, R. & Wanakule, N. 1993. A Lumped Parameter Model for the Edwards Aquifer. Texas Water Resources Institute.

Atkinson, T. C. 1977. Diffuse flow and conduit flow in limestone terrain in the Mendip Hills, Somerset (Great Britain). *Journal of Hydrology*, 35**,** 93-110.

Barrett, M. E. & Charbeneau, R. J. 1997. A parsimonious model for simulating flow in a karst aquifer. *Journal of Hydrology*, 196**,** 47-65.

Bennett, N. D., Croke, B. F. W., Guariso, G., Guillaume, J. H. A., Hamilton, S. H., Jakeman, A. J., Marsili-Libelli, S., Newham, L. T. H., Norton, J. P., Perrin, C., Pierce, S. A., Robson, B., Seppelt, R., Voinov, A. A., Fath, B. D. & Andreassian, V. 2013. Characterising performance of environmental models. *Environmental Modelling & Software*, 40**,** 1-20.

Beran, M. 1999. Hydrograph Prediction - How much skill? *Hydrology and Earth System Sciences*, 3**,** 305-307.

Beven, K. 2001. How far can we go in distributed hydrological modelling? *Hydrology and Earth System Sciences*, 5**,** 1-12.

Beven, K. & Binley, A. 1992. The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes*, 6**,** 279-298.

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.

Birtles, A. B. & Reeves, M. J. 1977. A simple effective method for the computer simulation of groundwater storage and its application in the design of water resource systems. *Journal of Hydrology*, 34**,** 77-96.

Bloomfield, J. 1997. The role of diagenesis in the hydrogeological stratification of carbonate aquifers: An example from the Chalk at Fair Cross, Berkshire, UK. *Hydrology and Earth System Sciences*, 1**,** 19-33.

Bloomfield, J. P., Bricker, S. H. & Newell, A. J. 2011. Some relationships between lithology, basin form and hydrology: a case study from the Thames basin, UK. *Hydrological Processes*, 25**,** 2518-2530.

Bloomfield, J. P., Gaus, I. & Wade, S. D. 2003. A method for investigating the potential impacts of climate-change scenarios on annual minimum groundwater levels. *Water and Environment Journal*, 17**,** 86-91.

Bloomfield, J. P. & Marchant, B. P. 2013. Analysis of groundwater drought using a variant of the Standardised Precipitation Index. *Hydrology and Earth System Sciences Discussions*, 10**,** 7537-7574.

Boorman, D. B., Hollis, J. M. & Lilly, A. 1995. Report No. 126 Hydrology of soil types: a hydrologically-based classification of the soils of the United Kingdom. Wallingford, UK: Institute of Hydrology.

Calver, A. 1997. Recharge Response Functions. *Hydrology and Earth System Sciences*, 1**,** 47-53.

Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., Wagener, T. & Hay, L. E. 2008. Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models. *Water Resources Research*, 44**,** W00B02.

Cleveland, R. B., Cleveland, W. S., McRae, J. E. & Terpenning, I. 1990. STL: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, 6**,** 3-73.

Conrads, P. A. & Roehl, E. A. 2007. Hydrologic Record Extension of Water-Level Data in the Everglades Depth Estimation Network (EDEN) Using Artificial Neural Network Models, 2000–2006. *U.S. Geological Survey Open-File Report 2007-1350.*

Coron, L., Andréassian, V., Perrin, C., Lerat, J., Vaze, J., Bourqui, M. & Hendrickx, F. 2012. Crash testing hydrological models in contrasted climate conditions: An experiment on 216 Australian catchments. *Water Resources Research*, 48**,** W05552.

Coulibaly, P., Anctil, F., Aravena, R. & Bobée, B. 2001. Artificial neural network modeling of water table depth fluctuations. *Water Resources Research*, 37**,** 885-896.

Cross, G. A., Rushton, K. R. & Tomlinson, L. M. 1995. The East Kent Chalk Aquifer during the 1988-92 Drought. *Water and Environment Journal*, 9**,** 37-48.

Daliakopoulos, I. N., Coulibaly, P. & Tsanis, I. K. 2005. Groundwater level forecasting using artificial neural networks. *Journal of Hydrology*, 309**,** 229-240.

Dawson, C. W. & Wilby, R. L. 2001. Hydrological modelling using artificial neural networks. *Progress in Physical Geography*, 25**,** 80-108.

Dooge, J. 1973. Linear Theory of Hydrologic Systems. Agricultural Research Service, U.S. Department of Agriculture, Washington, D.C.

Downing, R. A., Allender, R., Lovelock, P. E. R. & Bridge, L. R. 1970. The Hydrogeology of the River Trent River Basin. Institute of Geological Science, London, UK.

Eberts, S. M., Böhlke, J. K., Kauffman, L. J. & Jurgens, B. C. 2012. Comparison of particle-tracking and lumped-parameter age-distribution models for evaluating vulnerability of production wells to contamination. *Hydrogeology Journal*, 20**,** 263-282.

FAO, IIASA, ISRIC, ISSCAS & JRC 2012. Harmonized World Soil Database (version 1.2). FAO, Rome, Italy and IIASA, Laxenburg, Austria.

Field, M. 1983. The meteorological office rainfall and evaporation calculation system — MORECS. *Agricultural Water Management*, 6**,** 297-306.

Flores W, E. L., Gutjahr, A. L. & Gelhar, L. W. 1978. A stochastic model of the operation of a stream-aquifer system. *Water Resources Research*, 14**,** 30-38.

Gemitzi, A. & Stefanopoulos, K. 2011. Evaluation of the effects of climate and man intervention on ground waters and their dependent ecosystems using time series analysis. *Journal of Hydrology*, 403**,** 130-140.

Ghose, D. K., Panda, S. S. & Swain, P. C. 2010. Prediction of water table depth in western region, Orissa using BPNN and RBFN neural networks. *Journal of Hydrology*, 394**,** 296-304.

Goderniaux, P., Brouyère, S., Fowler, H. J., Blenkinsop, S., Therrien, R., Orban, P. & Dassargues, A. 2009. Large scale surface–subsurface hydrological model to assess climate change impacts on groundwater reserves. *Journal of Hydrology*, 373**,** 122-138.

Gossel, W., Ebraheem, A. M. & Wycisk, P. 2004. A very large scale GIS-based groundwater flow model for the Nubian sandstone aquifer in Eastern Sahara (Egypt, northern Sudan and eastern Libya). *Hydrogeology Journal*, 12**,** 698-713.

Griffiths, J., Keller, V., Morris, D. & Young, A. R. 2006. Continuous Estimation of River Flows (CERF) - Technical Report: Task 1.3: Model scheme for representing rainfall interception and soil moisture. Environment Agency R & D Project W6-101. Centre for Ecology and Hydrology, Wallingford, UK.

Howard, K. W. F. & Lloyd, J. W. 1979. The sensitivity of parameters in the Penman evaporation equations and direct recharge balance. *Journal of Hydrology*, 41**,** 329-344.

Ireson, A. M., Wheater, H. S., Butler, A. P., Mathias, S. A., Finch, J. & Cooper, J. D. 2006. Hydrological processes in the Chalk unsaturated zone – Insights from an intensive field monitoring programme. *Journal of Hydrology*, 330**,** 29-43.

Jackson, C. R., Meister, R. & Prudhomme, C. 2011. Modelling the effects of climate change and its uncertainty on UK Chalk groundwater resources from an ensemble of global climate model projections. *Journal of Hydrology*, 399**,** 12-28.

Jakeman, A. J. & Hornberger, G. M. 1993. How much complexity is warranted in a rainfall-runoff model? *Water Resources Research*, 29**,** 2637-2649.

Jakeman, A. J., Letcher, R. A. & Norton, J. P. 2006. Ten iterative steps in development and evaluation of environmental models. *Environmental Modelling & Software*, 21**,** 602-614.

Kazumba, S., Oron, G., Honjo, Y. & Kamiya, K. 2008. Lumped model for regional groundwater flow analysis. *Journal of Hydrology*, 359**,** 131-140.

Keating, T. 1982. A Lumped Parameter Model of a Chalk Aquifer-Stream System in Hampshire, United Kingdom. *Ground Water*, 20**,** 430-436.

Keller, V., Young, A. R., Morris, D. & Davies, H. 2006. Continuous Estimation of River Flows (CERF) - Technical Report: Task 1.1: Estimation of Precipitation Inputs. Environment Agency R & D Project W6-101. Centre for Ecology and Hydrology, Wallingford, UK.

Konikow, L. F. & Bredehoeft, J. D. 1992. Ground-water models cannot be validated. *Advances in Water Resources*, 15**,** 75-83.

Kooi, H. & Groen, J. 2003. Geological processes and the management of groundwater resources in coastal areas. *Netherlands Journal of Geosciences*, 82**,** 31-40.

Lees, M. J. 2000. Data-based mechanistic modelling and forecasting of hydrological systems. *Journal of Hydroinformatics*, 2**,** 15-34.

Legates, D. R. & McCabe, G. J. 1999. Evaluating the use of "goodness-of-fit" Measures in hydrologic and hydroclimatic model validation. *Water Resources Research*, 35**,** 233-241.

Maier, H. R. & Dandy, G. C. 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environmental Modelling & Software*, 15**,** 101-124.

Marsh, T. J. & Hannaford, J. 2008. UK Hydrometric Register. Centre for Ecology and Hydrology, Wallingford, UK.

Mathias, S. A., Butler, A. P., Jackson, B. M. & Wheater, H. S. 2006. Transient simulations of flow and transport in the Chalk unsaturated zone. *Journal of Hydrology*, 330**,** 10-28.

Mathias, S. A., Butler, A. P., McIntyre, N. & Wheater, H. S. 2005. The significance of flow in the matrix of the Chalk unsaturated zone. *Journal of Hydrology*, 310**,** 62-77.

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

Moore, R. J. & Bell, V. A. 2002. Incorporation of groundwater losses and well level data in rainfall-runoff models illustrated using the PDM. *Hydrology and Earth System Sciences*, 6**,** 25-38.

Nash, J. E. & Sutcliffe, J. V. 1970. River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, 10**,** 282-290.

Neuzil, C. E. 1986. Groundwater Flow in Low-Permeability Environments. *Water Resources Research*, 22**,** 1163-1195.

Olin, M. 1995. Estimation Of Base Level For An Aquifer From Recession Rates Of Groundwater Levels. *Hydrogeology Journal*, 3**,** 40-51.

Pappenberger, F., Beven, K., De Roo, A., Thielen, J. & Gouweleeuw, B. 2004. Uncertainty analysis of the rainfall runoff model LisFlood within the Generalized Likelihood Uncertainty Estimation (GLUE). *International Journal of River Basin Management*, 2**,** 123-133.

Park, E. & Parker, J. C. 2008. A simple model for water table fluctuations in response to precipitation. *Journal of Hydrology*, 356**,** 344-349.

Peck, A. 1988. Consequences of spatial variability in aquifer properties and data limitations for groundwater modelling practice. International Association of Hydrological Sciences, Oxfordshire, UK.

Pollock, D. W. 1994. User's guide for MODPATH/MODPATH-PLOT, Version 3; a particle tracking post-processing package for MODFLOW, the U.S. Geological Survey finite-difference ground-water flow model. *U.S. Geological Survey Open-File Report 94-464.*

Pozdniakov, S. P. & Shestakov, V. M. 1998. Analysis of groundwater discharge with a lumped-parameter model, using a case study from Tajikistan. *Hydrogeology Journal*, 6**,** 226-232.

Proctor, M. E., Siddons, P. A., Jones, R. J. A., Bellamy, P. H. & Keay, C. A. 1998. LandIS: the Land Information System for the UK. *In:* HEINEKE, H. J., ECKELMANN, W., THOMASSON, A. J., JONES, R. J. A., MONTANARELLA, L. & BUCKLEY, B. (eds.) *Land Information Systems: Developments for planning the sustainable use of land resources, EUR 17729 EN.* Office for Official Publications of the European Communities, Luxembourg.

Rushton, K. R. 2003. Groundwater hydrology: conceptual and computational models. John Wiley, Chichester, UK.

Rushton, K. R., Eilers, V. H. M. & Carter, R. C. 2006. Improved soil moisture balance methodology for recharge estimation. *Journal of Hydrology*, 318**,** 379-399.

Rushton, K. R. & Rao, S. V. R. 1988. Groundwater flow through a Miliolite limestone aquifer. *Hydrological Sciences Journal*, 33**,** 449-464.

Rushton, K. R. & Rathod, K. S. 1981. Aquifer response due to zones of higher permeability and storage coefficient. *Journal of Hydrology*, 50**,** 299-316.

Rushton, K. R., Smith, E. J. & Tomlinson, L. M. 1982. An improved understanding of flow in a limestone aquifer using field evidence and mathematical models. *Journal of the Institution of Water Engineers and Scientists*, 36.

Scanlon, B. R., Mace, R. E., Barrett, M. E. & Smith, B. 2003. Can we simulate regional groundwater flow in a karst system using equivalent porous media models? Case study, Barton Springs Edwards aquifer, USA. *Journal of Hydrology*, 276**,** 137-158.

Shepley, M. G., Whiteman, M. I., Hulme, P. J. & Grout, M. W. 2012. Groundwater Resources Modelling: A Case Study from the UK. Geological Society Special Publications 364, London, UK.

Smith, A. & Welsh, W. D. 2011. Review of groundwater models and modelling methodologies for the Great Artesian Basin. A technical report to the Australian Government from the CSIRO Great Artesian Basin Water Resource Assessment. *CSIRO Water for a Healthy Country Flagship Series.*

Sorensen, J. P. R., Finch, J. W., Ireson, A. M. & Jackson, C. R. 2014. Comparison of varied complexity models simulating recharge at the field scale. *Hydrological Processes*, 28**,** 2091-2102.

Sreekanth, P. D., Geethanjali, N., Sreedevi, P. D., Ahmed, S., Ravi Kumar, N. & Kamala Jayanthi, P. D. 2009. Forecasting groundwater level using artificial neural networks. *Current Science*, 96**,** 933-939.

Sun, D., Zhao, C., Wei, H. & Peng, D. 2011. Simulation of the relationship between land use and groundwater level in Tailan River basin, Xinjiang, China. *Quaternary International*, 244**,** 254-263.

Taylor, C. J. & Alley, W. M. 2001. Ground-Water-Level Monitoring and the Importance of Long-Term Water-Level Data. *U.S. Geological Circular 1217*.

Thiéry, D. 2012. Presentation of the GARDÉNIA v8.1 software package developed by BRGM. *BRGM Note technique (NT EAU 2012/01).*

Todd, D. K. 1959. Groundwater Hydrology. John Wiley, Chichester, UK.

Trichakis, I., Nikolos, I. & Karatzas, G. P. 2011. Comparison of bootstrap confidence intervals for an ANN model of a karstic aquifer response. *Hydrological Processes*, 25**,** 2827-2836.

Wada, Y., van Beek, L. P. H., van Kempen, C. M., Reckman, J. W. T. M., Vasak, S. & Bierkens, M. F. P. 2010. Global depletion of groundwater resources. *Geophysical Research Letters*, 37**,** L20402.

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.

Young, P. C., Castelletti, A. & Pianosi, F. 2007. The data-based mechanistic approach in hydrological modelling. *In:* CASTELLETTI, A. & SESSA, R. S. (eds.) *Topics on System Analysis and Integrated Water Resources Management.* Elsevier, Oxford, UK.

Young, P. C. & Lees, M. 1993. The Active Mixing Volume (AMV): a new concept in modelling environmental systems. *In:* BARNETT, V. & TURKMAN, K. F. (eds.) *Statistics for the Environment.* John Wiley, Chichester, UK.

8. Figures

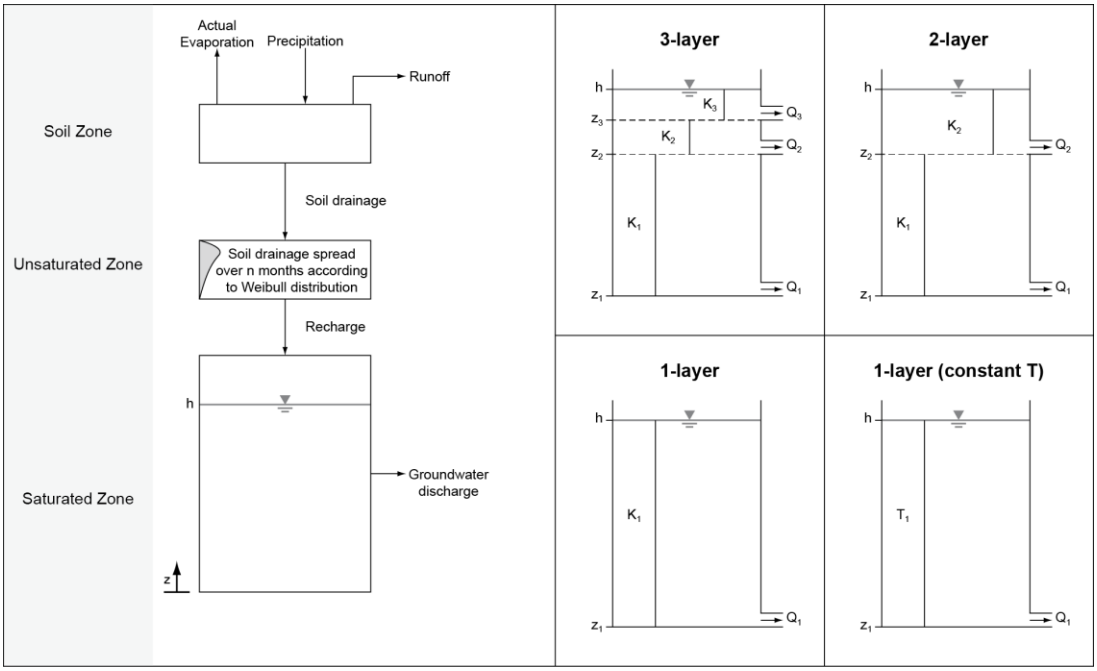

Figure 1: Schematic of generalised AquiMod model structure (left) and different saturated zone component structures used in this study (right).

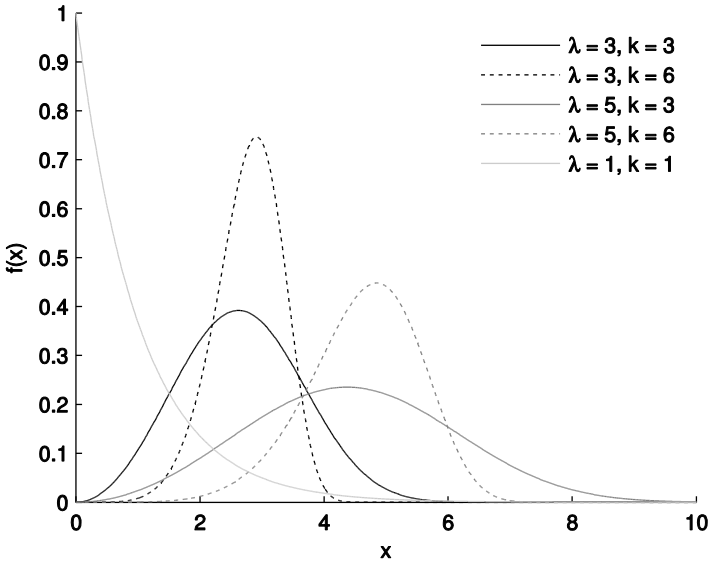

Figure 2: Probability distribution functions using the parametric Weibull distribution with different scale (λ) and shape (k) parameters.

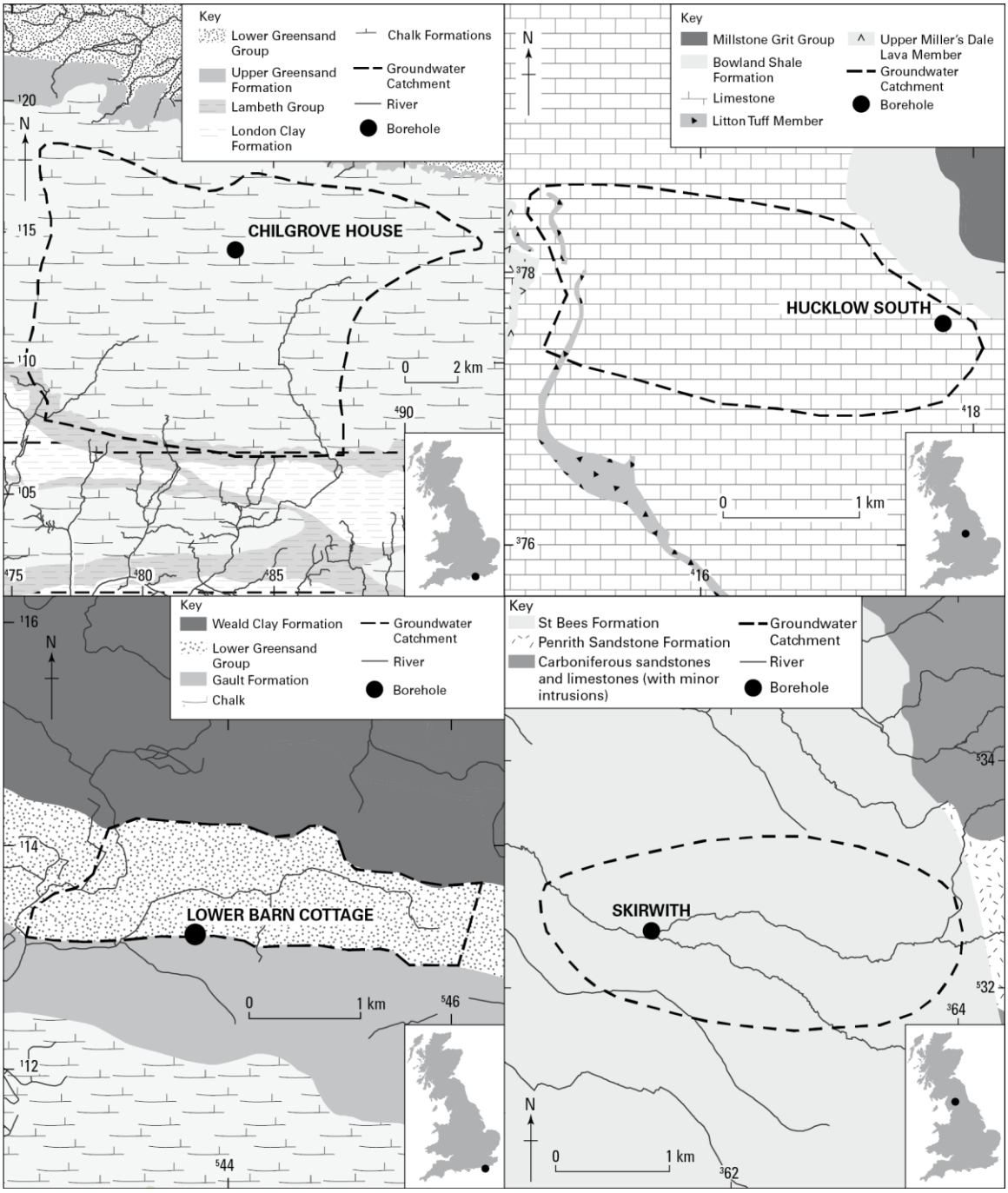

Figure 3: Location and geological setting of four observation boreholes.

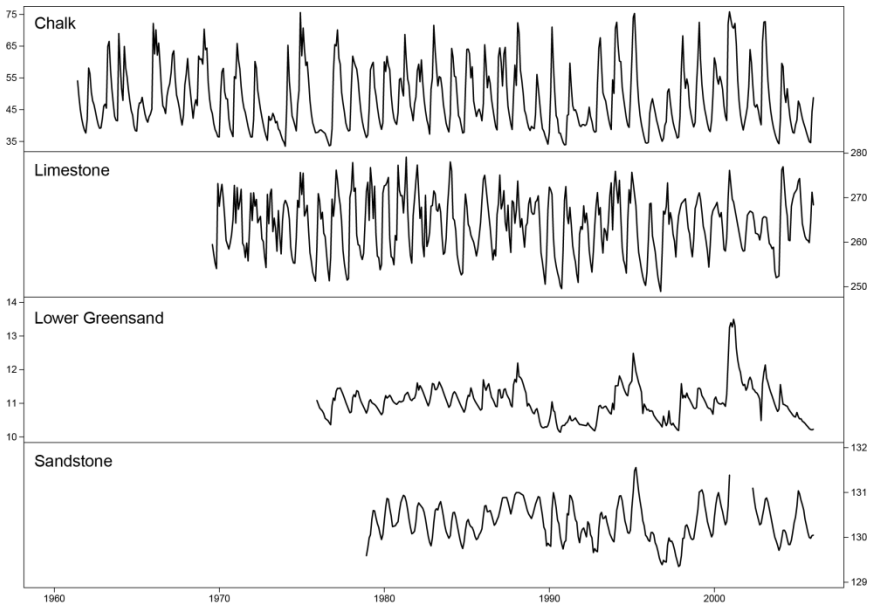

Figure 4: Observed groundwater level time-series in meters above sea level (m asl) at modelled sites.

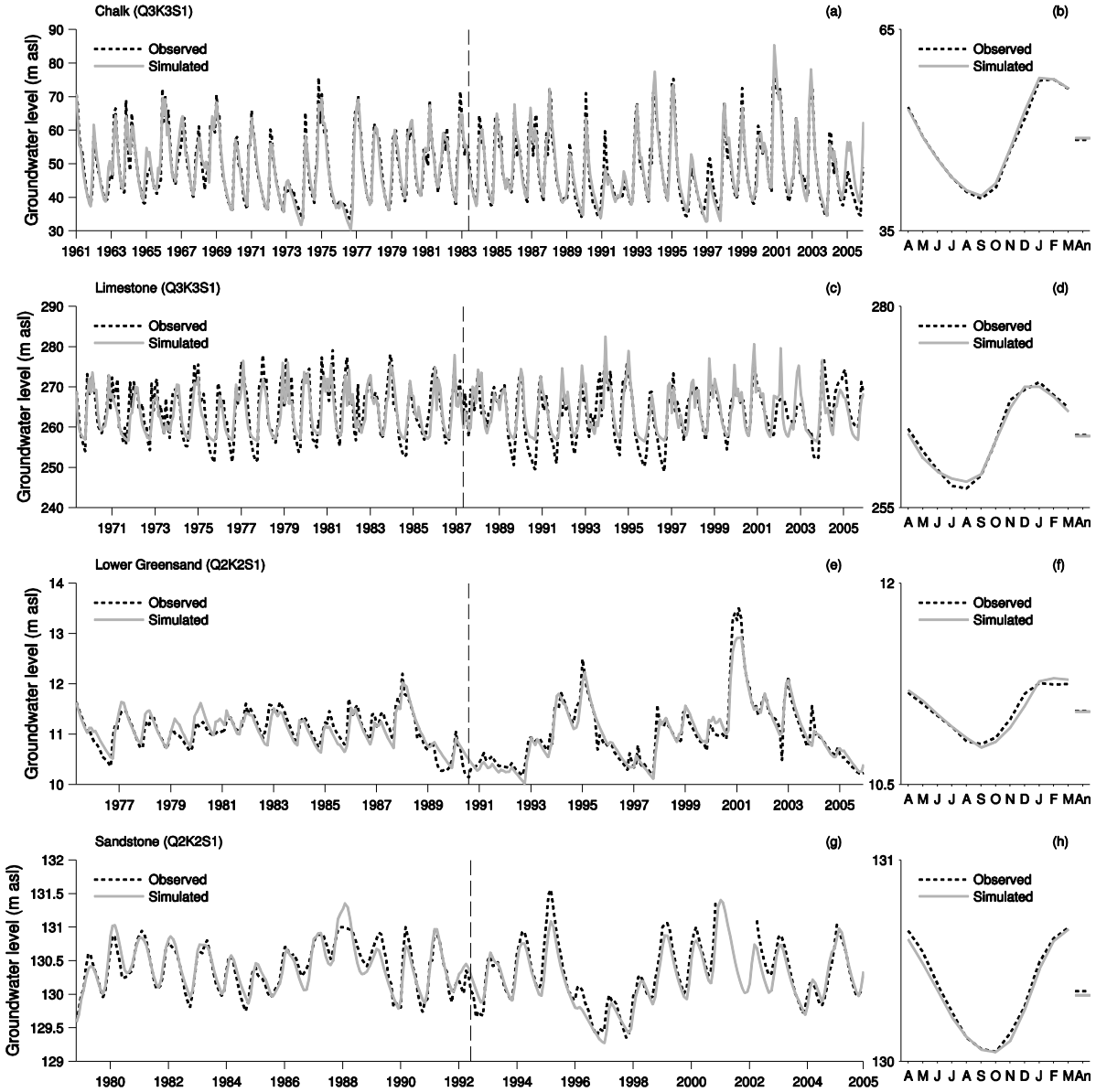

Figure 5: Observed groundwater level time-series and those simulated using the most efficient model structures (left), and mean monthly and annual groundwater levels (right).

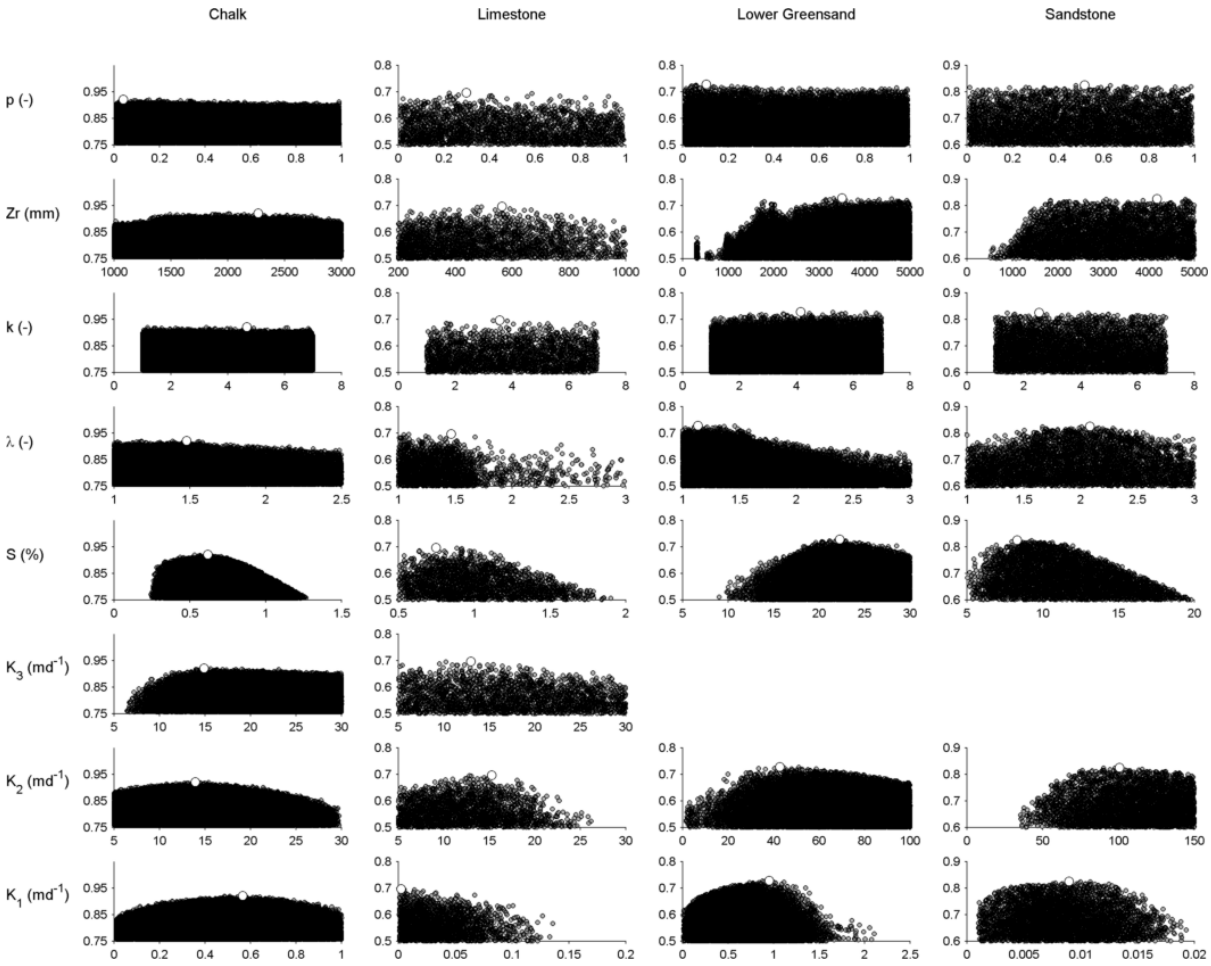

Figure 6: Dotty sensitivity plots for the most efficient model structures with parameter value on the x-axes and NSE on the y-axes. Black dots indicate individual parameters from Monte Carlo sampling and the white dot indicates the optimum parameter value.

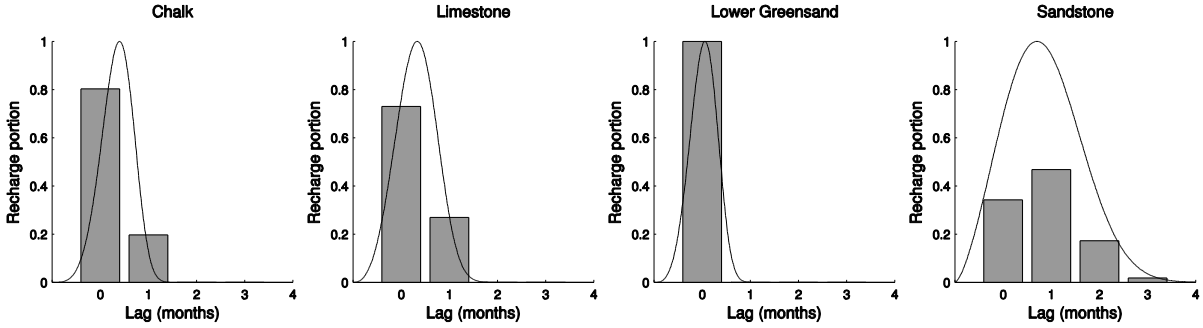

Figure 7: 1 Calibrated Weibull distribution (black line) and corresponding unsaturated zone function (grey bars) for the most efficient model structures.

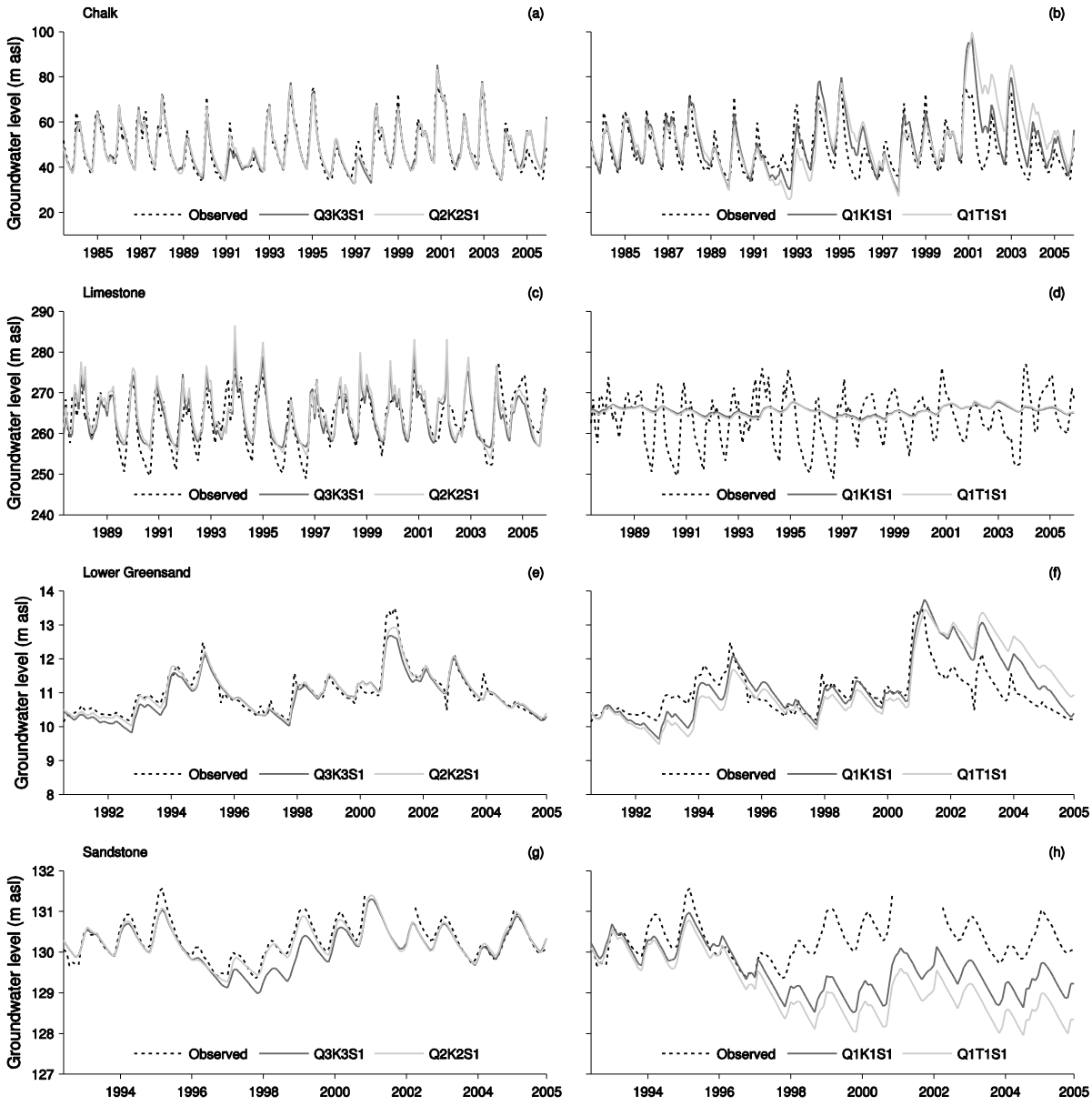

Figure 8: Observed and simulated groundwater level hydrographs over the evaluation sequence using four different groundwater model structures for the (a,b) Chalk, (c,d) Limestone, (e,f) Lower Greensand and (g,h) Sandstone study sites.

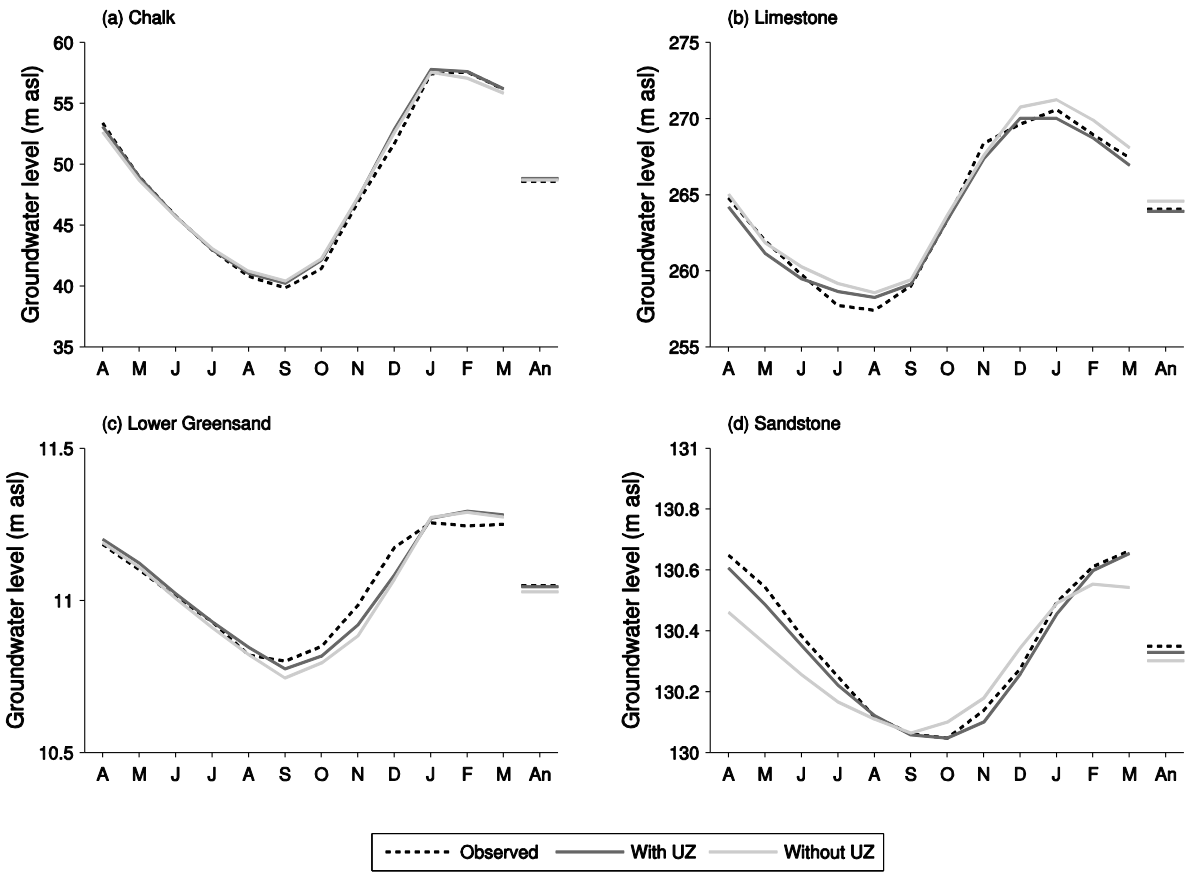

Figure 9: Observed and simulated mean monthly and annual groundwater levels with and without the unsaturated zone component.

9. Tables

Table 1: List of AquiMod model parameters arranged by module.

Module	**Parameter (units)**	**Description**
Soil	Δx (km)	Representative aquifer length
	BFI (-)	Baseflow index
	FC (-)	Field capacity of the soil
	WP (-)	Wilting point of the soil
	Zr (mm)	Maximum rooting depth of vegetation
	p (-)	Depletion factor of vegetation
Unsaturated Zone	n (-)	Maximum number of time-steps taken for soil drainage to reach the groundwater
	k (-)	Weibull shape parameter
	λ (-)	Weibull scale parameter
Saturated Zone	K_i (m d^{-1})	Hydraulic conductivity for layer i
	T_1 (m^2 d^{-1})	Transmissivity for layer 1 (one-layer fixed transmissivity component only)
	S (%)	Aquifer storage coefficient
	Z_i (m asl)	Outlet elevation for layer i

Table 2: Calibration (Cal), evaluation (Evl) and combined (Cmb) NSE and combined bias scores obtained for each of the five different model structures tested. The most efficient model structures based on their combined NSE scores are highlighted in grey.

	Chalk				Limestone				Greensand				Sandstone			
	Cal NSE	**Evl NSE**	**Cmb NSE**	**Cmb bias**	**Cal NSE**	**Evl NSE**	**Cmb NSE**	**Cmb bias**	**Cal NSE**	**Evl NSE**	**Cmb NSE**	**Cmb bias**	**Cal NSE**	**Evl NSE**	**Cmb NSE**	**Cmb bias**
3-layer	0.92	0.89	0.91	0.25	0.70	0.61	0.65	0.09	0.75	0.88	0.84	-0.04	0.83	0.52	0.66	-0.08
2-layer	0.91	0.89	0.90	0.09	0.69	0.54	0.62	0.97	0.73	0.92	0.88	0.00	0.82	0.86	0.85	-0.05
1-layer	0.73	0.49	0.60	1.57	0.00	0.02	-0.02	1.23	0.69	0.15	0.27	0.09	0.23	-4.23	-1.91	-0.40
1-layer (fixed T)	0.38	-0.28	0.02	1.56	-0.01	0.02	-0.01	1.03	0.67	-0.62	-0.33	0.09	0.30	-8.39	-4.73	-0.61
No UZ	0.91	0.89	0.90	0.13	0.71	0.56	0.64	0.55	0.74	0.91	0.87	-0.02	0.75	0.76	0.76	-0.05

Table 3: Optimum model parameter sets for the most efficient model structures. Calibration parameters highlighted in white and fixed parameters highlighted in grey. Prior calibration bands in square brackets.

Module	Parameter (units)	Chalk	Carboniferous Limestone	Lower Greensand	Triassic Sandstone
Soil Zone	Δx (km)	3.0	1.5	0.5	1.0
	BFI (-)	0.81	0.75	0.80	0.43
	FC (-)	0.290	0.285	0.286	0.286
	WP (-)	0.153	0.116	0.185	0.185
	Zr (mm)	2269 [1000-3000]	565 [200-1000]	3501 [1000-5000]	4180 [1000-5000]
	p (-)	0.04 [0-1]	0.30 [0-1]	0.1 [0-1]	0.52 [0-1]
Unsaturated Zone	n (-)	5	3	7	6
	k (-)	4.67 [1-7]	3.6 [1-7]	4.15 [1-7]	2.53 [1-7]
	λ (-)	1.47 [1-3]	1.47 [1-3]	1.13 [1-3]	2.08 [1-3]
Saturated Zone	K_3 (m d^{-1})	14.88 [5-30]	12.97 [5-30]	-	-
	K_2 (m d^{-1})	13.94 [5-30]	15.29 [5-30]	42.76 [1-100]	100.5 [10-150]
	K_1 (m d^{-1})	0.57 [0.01-1]	0.003 [0.001-0.2]	0.95 [0.01-5]	0.01 [0.001-0.05]
	S (%)	0.6 [0.1-1.5]	0.7 [0.5 – 5]	22.3 [1-30]	8.3 [5 – 20]
	z_3 (m asl)	49.8	262.6	-	-
	z_2 (m asl)	38.1	254.9	10.3	129.4
	z_1 (m asl)	-11.4	182.0	3.0	35.0