Robust evidence for random fractal scaling of ground water levels in unconfined aquifers

Max A. Little[1], John P. Bloomfield[2*]
[1]Systems Analysis, Modelling and Prediction Group, University of Oxford, UK
[2]British Geological Survey, Wallingford, Oxford, UK

Abstract

Temporal scaling in time series of groundwater level fluctuations has previously been quantified from a site on a loess/till aquifer from the USA. Using spectral analysis and detrended fluctuation (geometric) analysis techniques the scaling exponents were correlated with catchment parameters. This study introduces new approaches to improve the statistical robustness of such techniques for quantifying the fractal scaling of groundwater levels, and, for the first time, investigates scaling of groundwater levels from a relatively permeable aquifer. Six groundwater level time series and an associated river stage time series from the unconfined Chalk aquifer (a dual-porosity, fractured limestone aquifer) in the Pang-Lambourn catchment, UK, have been analysed. Surrogate data of time series with known scaling properties has been used to estimate the probability distribution of the spectral and geometric scaling exponents; robust regression techniques have been used to improve estimates of the scaling exponents; and robust non-parametric techniques have been used to correlate scaling exponents with features of the boreholes and catchments. Strong statistical support has been found for temporal scaling of groundwater levels over a wide range of time scales, however, bootstrap estimates of the scaling exponents indicate a much larger range of exponents than found by previous studies, suggesting that the uncertainty in existing estimates of scaling exponents may be too small. There is robust evidence that geometrical scaling properties at each borehole can be related to the depth of the observation boreholes and distance of those boreholes from the river in the catchment, but no such correlations were found for the spectral scaling exponents. The results build on the body of evidence that groundwater levels, as with many hydrogeological phenomena, may be well modeled with mathematical concepts from statistical mechanics that do not attempt to capture every detail of these highly heterogeneous and complex systems.

KEYWORDS: groundwater level; fractal scaling; spectral analysis; detrended fluctuation analysis;

Introduction

Understanding variations in groundwater levels is an essential element of water resource management and planning since groundwater levels represent the balance between two important catchment phenomena: recharge from precipitation, and groundwater discharge to rivers or streams. Unlike these two phenomena, which are difficult to monitor, groundwater level monitoring is relatively easy and groundwater levels, along with river levels and flow, provide a catchment-integrated indication of the status of water resources at the catchment scale.

It is well understood that mean groundwater levels are affected by the catchment water balance and bulk aquifer parameters, while individual excursions and seasonal variations in groundwater levels can be ascribed to individual rainfall events and seasonal variations in the driving variables. However, recharge and discharge phenomena act over a wide range of spatial (pore to catchment) and temporal (minutes to hundreds of years) scales, and are affected by a range of often highly non-linear processes and can be subject to feedbacks. They are influenced by highly heterogeneous porosity, saturation, and hydraulic conductivity fields found in aquifers, and are controlled by spatio-temporally varying driving variables, such as precipitation and evapo-transpiration. Consequently, groundwater levels never achieve a steady state and may vary over multiple spatial and temporal scales, and there is some recent evidence that groundwater levels may show scale-invariant, or fractal behaviour (Zhang and Schilling, 2004).

There is currently a major research effort amongst the catchment science community to seek to explain what have been referred to as *organising principles* (mathematically unifying relationships) that underlie catchment heterogeneity and complexity (McDonnell and Wood, 2004; Bevan, 2006; Sivapalan, 2005; McDonnell et al., 2007; Soulsby et al., 2008; Tetzlaff et al., 2008;). This has been cogently summarized by McDonnell et al. (2007) in a recent overview on the future aims of catchment science, as follows: “we need to move beyond the status quo of having to explicitly characterize or prescribe landscape heterogeneity … and in this way reproduce process complexity and instead explore the set of organising principles that might underlie the heterogeneity and complexity”. Fractal behaviour is a common feature of complex, heterogeneous natural systems and understanding what gives rise to fractal behaviour of groundwater levels should provide new, potentially significant insights into some of the organizing principles that may underlie the terrestrial water cycle. However, process-based studies, of the type that are commonly used in groundwater resource assessments, may not be appropriate in the development of models to predict the behaviour of complex systems and of fractal phenomena, and so other modelling strategies are required.

Typically, models of groundwater levels are based on conceptual *process models* which represent mechanisms associated with catchment discharge, recharge, saturated flow, baseflow and runoff. However, although process models have a long history and have proved to be invaluable for understanding the physical basis of groundwater flow dynamics, it is recognised that there are problems with such an approach. Prediction with all such *classical deterministic* process models is constrained by several mathematical limitations: (1) measurement error, nonlinearity and sensitivity to boundary conditions (chaos) (Smale, 1967), (2) model error (McSharry and Smith, 2004), and (3) inaccessible parameters and variables. Chaos occurs in many nonlinear systems when the temporal evolution of the model amplifies the error in the measurement of the boundary conditions: after a time, the state of the system becomes practically unpredictable, because the boundary conditions cannot be known to infinite precision. Model error occurs when the perfect model of the system is not known: it is usually the case that the model represents a simplification of a multitude of interacting, and often poorly-understood mechanisms. Finally, it is impossible to measure the parameters and variables of the aquifer at every spatial location – this poses a particular problem for the highly spatially heterogeneous nature of aquifers – exacerbating the uncertainty in predictions produced by the model.

Such problems with process models are not unique to hydrogeology: in meteorology for example, it has long been recognized that chaos and model error fundamentally limit prediction (Eady, 1951; Lorenz, 1963). The contemporary solution is essentially probabilistic: predictions are made that attempt to represent the full uncertainty due to chaos, produced by many randomized perturbations of the boundary conditions and model equations called *ensemble methods* (Buizza 2003). The successes of this approach have precipitated a major conceptual shift from deterministic to probabilistic modelling.

This shift may help to mitigate the mathematical limitations of process hydrogeological predictions, but it is not clear that this can also satisfactorily address the effect of high spatial heterogeneity coupled with inaccessible parameters, variables and boundary conditions (Beven, 2006). In practice, this may make it impossible to produce detailed predictions of groundwater levels with the same accuracy as, for example, daily surface temperatures. It may well be that the successes of groundwater level predictions resulting from process models calibrated against a few aquifer measurements could be the result of *overfitting*: that is, these predictions are accurate under limited conditions such as short time intervals or locations close to the borehole, but are erroneous for longer intervals or unmeasured sites.

A different, but useful, kind of statistical prediction may be possible with models rooted in the theory of *statistical mechanics*, as suggested by Eady (1951). These models have their origins as explanations for the observed bulk properties of gasses and fluids, where we are ignorant about the state variables of each particle, but precise statements can be derived about statistical properties of the model variables and derived quantities (Ruelle, 1984). This is similar to the situation with unmeasured variables and heterogeneous parameters in aquifers, and statistical mechanics models might therefore be co-opted to make predictions about the bulk properties of aquifers. Critically, these models use few parameters that must be inferred from measurements, significantly reducing the risk of overfitting.

Classical statistical mechanics explains the bulk statistical properties of simple systems such as ideal gasses. However, many, more complex, systems from diverse disciplinary origins show remarkably similar *scale-invariant* statistical fluctuations of their state variables. These fluctuations are *statistically self-similar* at all length scales, and this is one defining property of *stochastic fractals* (Falconer, 2003). Time series which have stochastic fractal noise, with power spectral density that scales as $f^{-\beta}$, where f is frequency and β is the *spectral scaling exponent*, have been observed from diverse disciplines. This has prompted theoretical explanations such as *self-organised criticality* (SOC) (Bak et al., 1988), *expansion-modification systems* (Li 1991), and *lattice gas density fluctuations* (Jensen, 1990). For example, SOC proposes that under constant small input flux, a local storage mechanism overflows into neighboring regions upon exceeding a capacity threshold. This situation causes cascading overflows on all length scales: time series from these simple models show scaling behaviour which is robust to variations in the model parameters. This suggests that this scaling behaviour is in some senses a universal property of complex media.

Since the pioneering work of Hurst (1951) on reservoir capacities, temporal and spatial scaling behaviour has been observed in time series and model outputs of many natural systems, including: earthquakes (Olami et al., 1992); fluvial and landscape evolution (Chase, 1992; Phillips, 2006; Murray and Fonstad, 2007); sandpiles (Bak et al., 1988); chemical reactions at mineral pore interfaces (Wells et al., 1991); rainfall (Lovejoy and Schertzer, 1985; Tessier et al., 1996); evapo-transpiration (Famiglietti et al., 2008); river water quality (Kircher et al., 2001); and runoff and river discharge (Pelletier and Turcotte, 1997; Kantelhardt et al., 2006; Koscielny-Bunde et al., 2006). To add to this list of scale invariant phenomena, Zhang and co-workers (Zhang and Schilling, 2004; Zhang and Li, 2005, 2006; Li and Zhang, 2007) have recently described scale invariance in groundwater levels from a single catchment on a till/loess system in the USA. These observations, along with the described mathematical limitations of classical process model predictions of groundwater levels and the utility of simple statistical mechanical models to explain scaling behaviour, are compelling arguments for the application of a statistical mechanical approach to the modelling of groundwater levels in permeable aquifers.

However, there remain many open questions. For example: how statistically reliable is the evidence supporting the scaling hypothesis for groundwater systems, and, how confident can we be about the typical range of scaling exponents? These questions must first be addressed before we can ask how these ranges of exponents relate to our current understanding of catchment characteristics, and what they tell us about any organizing principles that may control the scaling of groundwater levels. Unfortunately, answering these questions directly is complicated by the lack of theoretical understanding of the asymptotic statistical properties of the techniques (Mandelbrot and Wallis, 1969). This leaves residual doubts about the reliability of these findings which, in other contexts, have historically been subject to substantial revisions (Hamed, 2007).

Our main aim in this paper therefore is to provide more robust empirical evidence of scaling properties of groundwater levels backed up by extensive computation and two key statistical innovations: *surrogate data* and *robust regression*. Surrogate data are generated time series whose temporal scaling properties are known: synthesizing many of these time series allows bootstrap estimates of the distribution of scaling properties of the groundwater level time series under examination. Similarly, estimating temporal scaling properties requires straight-line regression of points on log-log scales, but classical least-squares regression is adversely affected by outliers, where robust regression is not. Using these innovations we explore computationally the statistical performance of spectral and geometric techniques for estimating temporal scaling exponents under known conditions. Having quantified this performance, we extend this to analysis of the unknown scaling properties of groundwater levels. Finally, we use robust non-parametric techniques to correlate these robustly estimated scaling exponents with features of the boreholes and their location in the catchment.

Methods

Our first task is to assess the evidence for scaling behaviour in borehole data. Firstly, we describe the classical formalism for stochastic fractal time series, which will allow analytical comparisons. We are interested in the class of time series, $x(t)$,that are Gaussian stochastic processes (that is, a set of Gaussian random variables indexed by the real time index t), with the property that $\mathrm{var}[x(t_1) - x(t_2)] \propto |t_1 - t_2|^{2H}$ for arbitrary time indices t_1, t_2. This condition implies that $x(t)$ and $s^{-H} x(st)$ have the same distribution, for all scale factors $s > 0$. The parameter H is the *scaling exponent* (also known as the *Hurst exponent*) of the self-similar process. Informally, as H increases, the resulting stochastic time series becomes smoother. Since the *autocorrelation* can be calculated directly from this definition, it is also straightforward to show (Falconer, 2003) that the power spectrum $X(f) = f^{\beta}$, where the spectral scaling exponent $\beta = 2H + 1$. The measurements in this study are available at fixed time intervals, where $t_n = n \bullet \Delta t$, i.e. we have $x_n = x(t_n)$. We can simulate approximately self-similar Gaussian time series at these time points t_n using the inverse discrete-time Fourier transform (hereafter, this is referred to as the *power spectral method*, PSM). Furthermore, we can estimate the spectral scaling exponent using the forward discrete-time Fourier transform, and since $-\log X(f)/\log f = \beta$, the slope of the log-log plot of f against $X(f)$ is an estimate of the spectral scaling exponent. We call this the power spectral density (PSD) scaling exponent estimation method.

The statistical self-similarity of the time series suggests an alternative formalism related to *broken-line processes* (Bras and Rodríguez-Iturbe, 1985). The *random midpoint displacement* (RMD) algorithm can simulate approximately self-similar Gaussian time series on t_n . For a time series of length N that is a power of two, it involves successive subdivision in stages numbered $k = 1, 2 \ldots \log_2 N$, and in the first stage the midpoint is set to $x_{N/2} = 1/2\,(x_N + x_1) + \varepsilon$, where ε is a Gaussian random variable of zero mean. We then linearly interpolate the time points between $[1, N/2]$ and between $[N/2, N]$. The next stage, $k = 2$, sets the new midpoints $N/4$, $3N/4$ according to the same random midpoint displacement scheme. This process repeats until all time points are calculated. The variance of ε at each stage is set to $[1 - 2^{2H-2}] / 2^{2kH}$.

Similarly, the scaling exponent of self-similar time series can also be estimated using successive subdivision. By definition, the standard deviation (*fluctuation*) over any sub-interval of length L of the Gaussian time series, will be approximately L^H (Falconer, 2003). Therefore, we can estimate H by first dividing up the time series into sub-intervals of length L, estimating the variance of each sub-interval, and averaging over each standard deviation estimate. Then, by increasing L and repeating the standard deviation calculations over this new sub-interval size, we can estimate H by the slope of the log-log plot of L against the average standard deviation of sub-intervals at each L. *Detrended fluctuation analysis* (DFA) proposes two advances over this basic algorithm. Firstly, although self-similar time series are essentially unbounded (as the variance increases with L), groundwater level time series are bounded, so that estimates of the larger scales are poor. By integrating the time series with the mean removed, i.e. by calculating $\hat{x}_n = \sum_{i=1}^{n} \left(x_i - E[x]\right)$, estimates of the scaling at larger sub-intervals are improved. Secondly, most groundwater level time series have trends and other local variations due to factors such as climate variation. By removing local linear trends in each sub-interval (by fitting a straight line or higher-order polynomial to the integrated time series $\hat{x}_n$ and subtracting this), estimates of H robust to these trends can be obtained. It can be shown that the slope α of the log-log plot of L against the average standard deviation $F(L)$ is equal to $H - 1$, which is the effect of integrating to obtain $\hat{x}_n$.

Here we introduce an innovation to obtain more robust estimates of groundwater level scaling exponents α and β. Reliable scaling exponent estimates generally require that the log-log plots lie on a straight line (are collinear) over a very large range of length scales. This is often difficult to obtain in practice, because most groundwater level time series are short or have measurement error that may well be temporally correlated. Either the smallest or largest scales will be unusable, or there may be length scales that are outliers, in the sense that although most of the points are collinear, a few points are not, and we wish to discard these when using line fitting to estimate the slope. We use *iteratively reweighted least squares* line-fitting with *Huber penalty function* (Hastie et al., 2001), which concentrates the slope estimate only on those points that are most collinear. It should be noted that this effectively circumvents analysis of *crossovers* – potential changes in scaling properties at different time scales – but gives more realistic estimates of the overall scaling behaviour, which is the main aim of this paper.

The performance of the PSD and DFA methods on PSM and RMD bootstrap time series are shown in Figure 1. Minimum/maximum values were assessed by generating 100 fractal time series with the same algorithm. As expected, the PSD method performs almost perfectly on PSM noise, because the method of generating the noise and measuring its scaling exponent are essentially the same. On RMD noise, however, the PSD method performs quite poorly for exponents $\beta < 1$ and $\beta > 2$. The DFA method performs well for PSM time series with exponents $\alpha < 1.2$, but otherwise, it shows a significant deviation away from the true value, although the deviation is not as severe as with PSD on RMD noise. Finally, the DFA method performs very well for RMD noise for $\alpha > 0.8$; for $\alpha < 0.8$, there is a significant deviation away from the true value, but again not as severe as with PSD applied to RMD noise. These findings suggest that, except for the PSD method on RMD noise with high β, although an exact value for the scaling exponent is not always possible, the estimated scaling value always increases with the true value, such that comparisons between estimated values are always indicative of a comparison between the underlying, true values.

Data

The groundwater level and river stage data used in this study, come from a research site at Boxford, Berkshire, UK, Figure 2. The study site has been previously described by (Gooddy et al., 2006), but is summarized here. It is centered on the River Lambourn, a rural, predominantly groundwater-fed catchment (~200km^2, Baseflow Index 0.96, mean flow ~1.75m^{3}sec^{-1}) which drains part of the Chalk aquifer of the Berkshire Downs. The site is underlain by thin soils, typically <1m thick. Alluvial sands and gravels are present adjacent to and below the river to a depth of about 3m, these in turn overlie up to 200m of Chalk. The Chalk is the main regional aquifer in the UK, with a mean matrix porosity of 39%, mean storage coefficient of 0.006, and transmissivity in the range 0.5 to ~8000 m^{2}d^{-1} with a geometric mean of 620 m^{2}d^{-1} (Bloomfield et al., 1995; Allen et al., 1997).

Groundwater levels have been monitored at six locations at the site and the river level has been monitored using a stilling well for up to five years, Table 1. Water levels at the monitoring locations were measured using pressure transducers and data loggers with a measuring range of 10mH_2O and a measurement resolution of 0.2cmH_2O. The sampling rate was either hourly or at 15-minute intervals. The resulting time series lengths varied from N = 19,750 to N = 49,133, Table 1. Each of the seven water level time series has been normalized to the range [-1, 1] for subsequent analysis, Figure 3.

Results

Having described the techniques, we now apply these to the normalized water level time series. Figure 4 shows log-log plots of DFA sub-interval size L against fluctuation $F(L)$, and frequency bin i against power spectral amplitude $|X(i)|^2$ of the time series x_n. It also shows the estimated scaling exponents α and β for each time series. The DFA sub-intervals ranged on a logarithmic scale from $L = 4$ to $L = N/2$ points. The PSD frequency bins ranged from $i = 2$ to $i = N/20$, because the power spectral scaling did not extend past this range of frequencies.

Figure 4 shows that, over the range of time scales where scaling behaviour could be reliably estimated, the data can be well modeled by a random fractal stochastic process, both in terms of spectral (PSD) and geometric (DFA) scaling. In order to assess whether there is any statistically significant difference between these exponents on the different water level time series, we generated a new set of 20 realizations of stochastic processes with the same scaling exponents as estimated from the data. Since the DFA method is most reliable on RMD noise, we used RMD realizations for α estimates, and for the same reason, for the PSD method used PSM realizations for the β estimates. Figure 5 shows the result, where the distributions are obtained using Gaussian kernel density estimation. This shows that the distributions of the scaling exponents are clearly quite different for each water level time series, for both spectral and geometric exponents. Using the non-parametric Kolmogorov-Smirnov test, we find that all distributions are significantly different ($p < 0.05$, $n = 20$) for all pairwise combinations of water level series. The obtained values of the scaling exponents are summarized in Table 2.

We also assessed the extent of (non-parametric) correlation between selected geometric properties of the borehole and the fractal scaling exponents (see Table 3). This shows that although the spectral scaling β is not significantly correlated with the distance of the observation point from either the river or the stilling well in the river (site PL26U), or with the mean observation depth, the geometric scaling exponent α shows large correlations with all these parameters.

Discussion

In this study, we assessed the evidence for random fractal scaling behaviour in groundwater level time series, towards providing evidence on which to advance statistical mechanical models of the dynamics of unconfined aquifers. Having noted the connection between statistical mechanical models and self-similar time series, we formally defined spectrally scaled Gaussian stochastic and statistically self-similar time series. We then described two methods for generating such time series with given scaling exponents, and rehearsed two complementary methods for estimating the scaling exponents from time series. Using innovations to improve the robustness of these estimation techniques, we applied them to water level time series from an unconfined aquifer and found scaling behaviour over a wide range of time scales. Using nonparametric techniques, we found robust statistical evidence that different groundwater level series exhibit different scaling properties. We also find evidence that the geometric scaling properties at each borehole are related to the basic physical layout of the aquifer, in particular to the distance from the river and depth of the observation zone. However, we found that the spectral scaling properties of the time series were unrelated to aspects of the physical layout of the aquifer that we tested.

These findings build on the growing body of evidence that supports the scaling hypothesis in groundwater levels (Zhang and Schilling, 2004; Li and Zhang, 2007), and extends the observation to more permeable aquifers than previously reported. The bootstrap estimates of the geometric (DFA) scaling exponent range from around 1.20 to 1.65 (Figure 5), which agrees approximately with the range found by Li and Zhang (2007), i.e. 1.28 to 1.64. However, bootstraps lead to much larger ranges of scaling exponents than those found previously – our suggestion is that the uncertainty in existing estimates is too small.

Also in agreement with Li and Zhang (2007), we found that the geometrical 'roughness' of the time series decreases with increasing distance from one of the external driving sources (here, the river flow), which is physically intuitive because the aquifer is a storage medium that tends to 'dampen' short-time variations in driving variables. We quantified this relationship as being particularly strong, $\rho > 0.8$. A novel finding is that this relationship is not detected in spectral scaling exponents. One plausible explanation is that this is a consequence of the limitations of classical linear analysis that measures only statistical means and covariances. The evidence suggests that classical linear spectral analysis does not extract sufficient information from groundwater levels to detect this physical phenomenon.

Our methods were designed exclusively to improve the robustness of the evidence of the basic scaling hypothesis in groundwater levels, so we cannot compare existing crossover findings (Li and Zhang, 2007) with our results. However, a natural extension of this study would be to devise similarly robust methods across limited ranges of time scales, and also of interest in future work would be the investigation of *multifractal* scaling (Kantelhardt et al., 2006).

Conclusions

Our main conclusion is that these results provide a sound statistical basis for supporting the investigation of simple statistical mechanical models as *highly parsimonious* dynamical explanations for the behaviour of groundwater levels. Classical linear models for groundwater flow would be unable to parsimoniously represent fractal scaling – a statistical mechanical model may actually be simpler, because classical linear systems require *infinite memory* to replicate the self-similar behaviour of the measured groundwater levels, whereas nonlinear models require only finite memory. More details of this line of reasoning can be found in Bras et al. (1985).

We hope that these findings motivate further research into statistical mechanical modelling of such systems, as a complementary approach to classical process-based modelling. For example, there is the need to discover dynamical explanations for these findings, in terms of parameter ranges and simple statistical state transition rules. Also needed is a comparison of these results against simulations from existing numerical groundwater models.

These results suggest that an explanation for the scale invariance of groundwater levels in unconfined aquifers as a 'complex' response to constantly changing driving inputs and boundary conditions (including boundaries imposed by management regimes) should be considered. These observations should provide additional impetus to the search for underlying organizing principles that may relate the scaling characteristics of recharge, groundwater head and discharge in permeable catchments.

Acknowledgements

Some of the data described was obtained as part of the Natural Environment Research Council's Lowland Permeable Catchment Research (LOCAR) Programme (http://catchments.nerc.ac.uk/). John Bloomfield publishes with the permission of the Executive Director of the British Geological Survey (NERC).

References

Allen DJ, Brewerton LJ, Coleby LM,Gibbs BR, Lewis MA, MacDonald AM, Wagstaff SJ, Williams AT. 1997. *The physical properties of major aquifers of England and Wales*. British Geological Survey, Technical Report, WD/97/34.

Bak, P., C. Tang and K. Wiesenfeld. 1988. Self-organized criticality. *Physical Review A,* **38,** 364 - 373.

Beven, K. 2006. Searching for the Holy Grail of scientific hydrology: Q(t) = H(S,R,Dt)A as closure. *Hydrology and Earth System Sciences,* **10,** 609-618.

Bloomfield JP, Brewerton LJ. Allen DJ. 1995. Regional trends in matrix porosity and dry density of the Chalk of England. *Quarterly Journal of Engineering Geology,* **28,** S131-S142

Bras, R. L. and I. Rodríguez-Iturbe. 1985. *Random functions and hydrology*. Reading, Mass., Addison-Wesley.

Buizza, R. 2003. Weather prediction: Ensemble prediction. *Encyclopaedia of Atmospheric Sciences*. J. R. Holton, J. Pyle and J. A. Curry. London, Academic Press.

Chase, C. G. 1992. Fluvial Landsculpting and the Fractal Dimension of Topography. *Geomorphology,* **5,** 39-57.

Eady, E. T. 1951. The Quantitative Theory of Cyclone Development. In: *Compendium of Meteorology*. T. F. Malone (Ed.), American Meteorological Society**,** pages 464-469.

Falconer, K. J. 2003. *Fractal geometry : mathematical foundations and applications*. Chichester, England, Wiley.

Famiglietti, J. S., D. R. Ryu, A. A. Berg, M. Rodell and T. J. Jackson. 2008. Field observations of soil moisture variability across scales. *Water Resources Research,* **44,** W01423, doi:10.1029/2006WR005804.

Gooddy, D. C., W. G. Darling, C. Abesser and D. J. Lapworth. 2006. Using chlorofluorocarbons (CFCs) and sulphur hexafluoride (SF6) to characterise groundwater movement and residence time in a lowland Chalk catchment. *Journal of Hydrology,* **330,** 44-52.

Hamed, K. H. 2007. Improved finite-sample Hurst exponent estimates using rescaled range analysis. *Water Resources Research* **43,** W04413, doi:10.1029/2006WR005111.

Hastie, T., R. Tibshirani and J. H. Friedman. 2001. *The elements of statistical learning : data mining, inference, and prediction : with 200 full-color illustrations*. New York, Springer.

Hurst, H. E. 1951. Long-Term Storage Capacity of Reservoirs. *Transactions of the American Society of Civil Engineers,* **116,** 770-799.

Jensen, H. J. 1990. Lattice Gas as a Model of 1/F Noise. *Physical Review Letters,* **64,** 3103-3106.

Kantelhardt, J. W., E. Koscielny-Bunde, D. Rybski, P. Braun, A. Bunde and S. Havlin. 2006. Long-term persistence and multifractality of precipitation and river runoff records. *Journal of Geophysical Research-Atmospheres,* **111,** D01106**,** doi:10.1029/2005JD005881.

Kirchner J.W., X. Feng, and C. Neal. 2001. Catchment-scale advection and dispersion as a mechanism for fractal scaling in stream tracer concentrations. Journal of Hydrology, 254, 82-101.

Koscielny-Bunde, E., J. W. Kantelhardt, P. Braun, A. Bunde and S. Havlin. 2006. Long-term persistence and multifractality of river runoff records: Detrended fluctuation studies. *Journal of Hydrology,* **322,** 120-137.

Li, W. T. 1991. Expansion-Modification Systems - a Model for Spatial 1/F Spectra. *Physical Review A,* **43,** 5240-5260.

Li, Z. W. and Y. K. Zhang. 2007. Quantifying fractal dynamics of groundwater systems with detrended fluctuation analysis. *Journal of Hydrology,* **336,** 139-146.

Lorenz, E. N. 1963. Deterministic Nonperiodic Flow. *Journal of the Atmospheric Sciences,* **20,** 130-141.

Lovejoy, S. and D. Schertzer. 1985. Generalized Scale-Invariance in the Atmosphere and Fractal Models of Rain. *Water Resources Research,* **21,** 1233-1250.

Mandelbrot, B. B. and J. R. Wallis. 1969. Robustness of Rescaled Range R/S in Measurement of Noncyclic Long Run Statistical Dependence. *Water Resources Research,* **5,** 967-988.

McDonnell, J.J and R. Wood. 2004. On the need for catchment classification. *Journal of Hydrology,* **299**, 2-3.

McDonnell, J. J., M. Sivapalan, K. Vache, S. Dunn, G. Grant, R. Haggerty, C. Hinz, R. Hooper, J. Kirchner, M. L. Roderick, J. Selker and M. Weiler. 2007. Moving beyond heterogeneity and process complexity: A new vision for watershed hydrology. *Water Resources Research,* **43,** W07301, doi:10.1029/2006WR005476**.**

McSharry, P. E. and L. A. Smith. 2004. Consistent nonlinear dynamics: identifying model inadequacy. *Physica D-Nonlinear Phenomena,* **192,** 1-22.

Murray, B. and M. A. Fonstad. 2007. Preface: Complexity (and simplicity) in landscapes., *Geomorphology,* **91,** 173-177.

Olami, Z., H. J. S. Feder and K. Christensen. 1992. Self-organized criticality in a continuous, nonconservative cellular automaton modeling earthquakes. *Physical Review Letters,* **68,** 1244-1247.

Pelletier, J. D. and D. L. Turcotte. 1997. Long-range persistence in climatological and hydrological time series: analysis, modeling and application to drought hazard assessment. *Journal of Hydrology,* **203,** 198-208.

Phillips J., D. 2006. Evolutionary geomorphology: thresholds and nonlinearity in landform response to environmental change. *Hydrology and Earth System Science*, **10**, 731-742.
Ruelle, D. 1984. *Thermodynamic formalism : the mathematical structures of classical equilibrium statistical mechanics*. New York, NY, USA, Cambridge University Press.
Sivapalan, M. 2005. Pattern, process and function: Elements of a new unified hydrologic theory at the catchment scale. In: Encyclopaedia of Hydrologic Sciences, 1(1) (Ed. M.G. Anderson), Chap. 13, pages 193-219. John Wiley, Hobokem, N.J.
Smale, S. 1967. Differentiable Dynamical Systems. I. Diffeomorphisms. *Bulletin of the American Mathematical Society*, **73,** 747-817.
Soulsby, C. C. Neal, H. Laudon, D.A> Burns, P. Merot, M. Bonel, S.M. Dunn, and D. Tetzlaff. 2008. Catchment data for process conceptualisation: simply not enough? Hydrological Processes, **22**, 2057-2061.
Tessier, Y., S. Lovejoy, P. Hubert, D. Schertzer and S. Pecknold. 1996. Multifractal analysis and modeling of rainfall and river flows and scaling, causal transfer functions. *Journal of Geophysical Research-Atmospheres*, **101,** 26427-26440.
Tetzlaff, D., J. J. McDonnell, S. Uhlenbrook, K. J. McGuire, P. W. Bogaart, F. Naef, A. J. Baird, S. M. Dunn and C. Soulsby. 2008. Conceptualizing catchment processes: simply too complex? *Hydrological Processes*, **22,** 1727-1730.
Wells, J. T., D. R. Janecky and B. J. Travis. 1991. A Lattice Gas Automata Model for Heterogeneous Chemical-Reactions at Mineral Surfaces and in Pore Networks. *Physica D* **47,** 115-123.
Zhang, Y. K. and Z. W. Li. 2005. Temporal scaling of hydraulic head fluctuations: Nonstationary spectral analyses and numerical simulations. *Water Resources Research*, **41,** W07031 , 10.1029/2004WR003797.
Zhang, Y. K. and Z. W. Li. 2006. Effect of temporally correlated recharge on fluctuations of groundwater levels. *Water Resources Research*, **42,** W10412, 10.1029/2005WR004828.
Zhang, Y. K. and K. Schilling. 2004. Temporal scaling of hydraulic head and river base flow and its implication for groundwater recharge. *Water Resources Research*, **40,** W03504, 10.1029/2003WR002094.

Table 1: Tabulated description of water level data from the Boxford site used in this study.

Borehole ID (measurement type)	Easting	Northing	Distance from PL26U (m)	Distance from river (m)	Observation zone depth (m bGL)	Geology	Land cover	Sample rate (mins)	Record start	Record end	Missing data
PL26E	442804	172269	16.5	16.0	18.0	Chalk	Woodland	60	23/12/02	05/03/08	Seven 1hr gaps; one 27hr gap.
PL26F	442800	172232	53.0	53.0	22.4	Chalk	Woodland	60	23/12/02	26/08/05	Four 1hr gaps.
PL26G	442829	172478	195.2	193.0	63.8	Chalk	Arable	60	05/03/04	06/06/06	One 2 hour gap.
PL26H	442814	172340	56.8	55.0	27.5	Chalk	Arable	60	10/01/03	06/04/06	One 1 hr gap.
PL26I	442822	172409	126.0	124.0	45.9	Chalk	Arable	60	23/12/02	03/03/08	
PL26Q	442834	172292	34.7	7.0	2.0	Gravel	Arable	15	22/02/07	18/07/08	Six gaps 1hr to 6hrs.
PL26U (stilling well - river)	442800	172285	0.0	0.0	0.0	River	Water	15	22/02/07	28/10/08	

Table 2. Stochastic fractal scaling exponents α, β obtained for water level time series from the Boxford site. The confidence intervals are one standard deviation, estimated using 20 realizations of stochastic processes with the same scaling exponents as estimated from the data.

	PL26E	*PL26I*	*PL26Q*	*PL26U*	*PL26G*	*PL26H*	*PL26F*
β	2.00±0.02	2.43±0.02	2.62±0.02	1.94±0.03	2.08±0.04	1.98±0.04	1.65±0.05
α	1.43±0.03	1.49±0.04	1.40±0.03	1.29±0.02	1.53±0.03	1.52±0.03	1.30±0.02

Table 3. Spearman rank correlation coefficients ρ of stochastic fractal scaling exponents α, β against three parameters related to the Boxford site. Entries marked '*' are significant at the 95% confidence level.

	Distance from PL26U against α	*Distance from river against α*	*Depth of observation zone against α*	*Distance from PL26U against β*	*Distance from river against β*	*Depth of observation zone against β*
Correlation ρ	0.8214*	0.8571*	0.8571*	0.3214	0.2143	0.2143
Correlation p-value	0.0341	0.0238	0.0238	0.4976	0.6615	0.6615

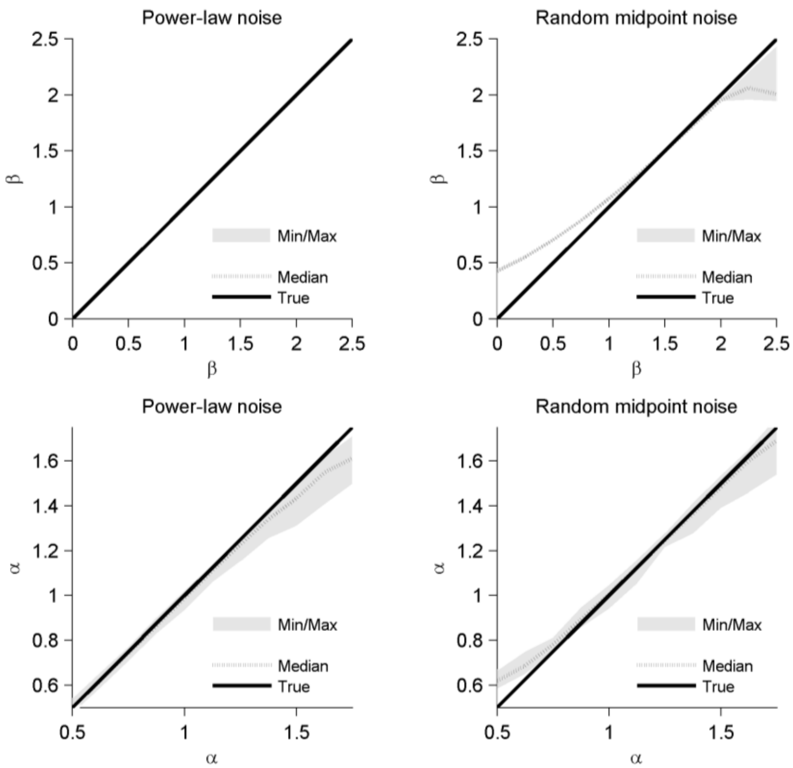

Figure 1: Performance of power-spectral density (estimating β) and detrended fluctuation analysis (estimating α) methods on power-spectral (PSM) and random midpoint (RMD) noise.

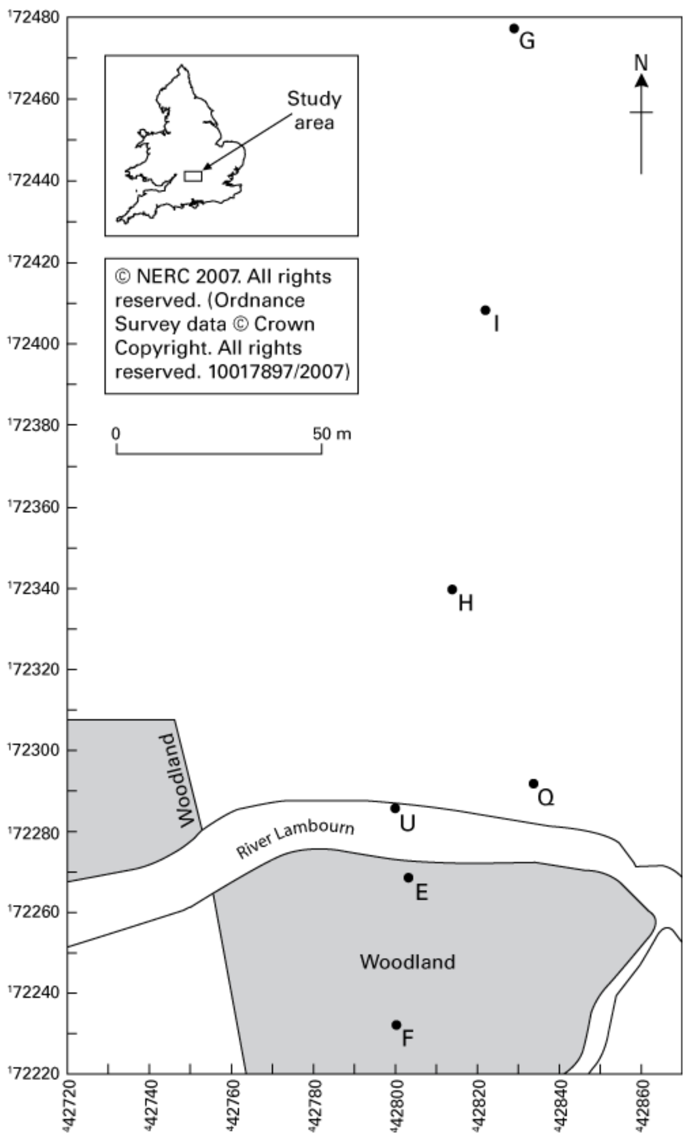

Figure 2: Schematic illustration of the Boxford site showing the relative locations of the boreholes and the stilling well in the river Lambourn (PL26U).

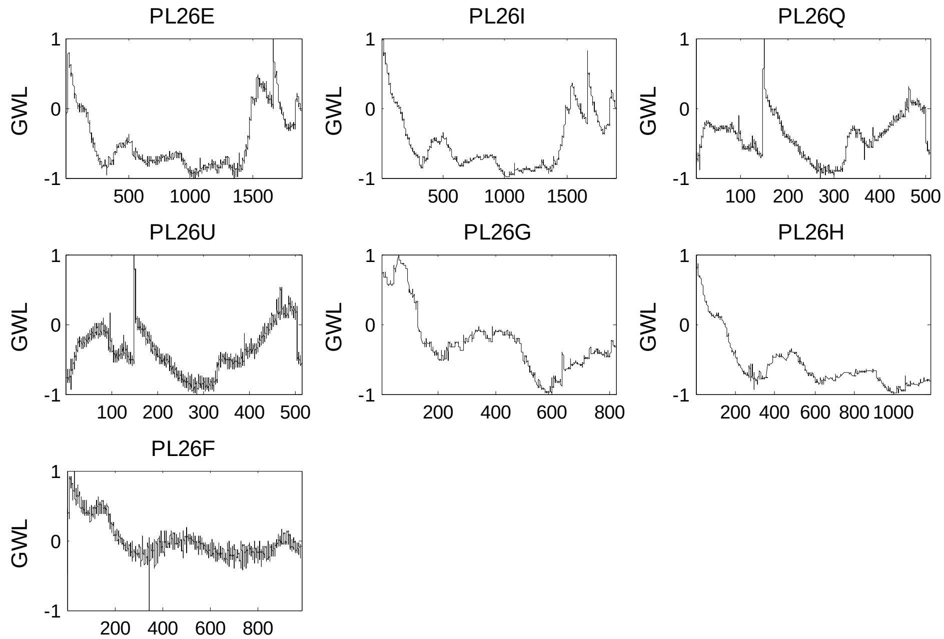

Figure 3: Normalised water level (NWL) time series from the Boxford site. The vertical axis is unitless, the horizontal axis is time in days since the start of the record, excluding missing measurements.

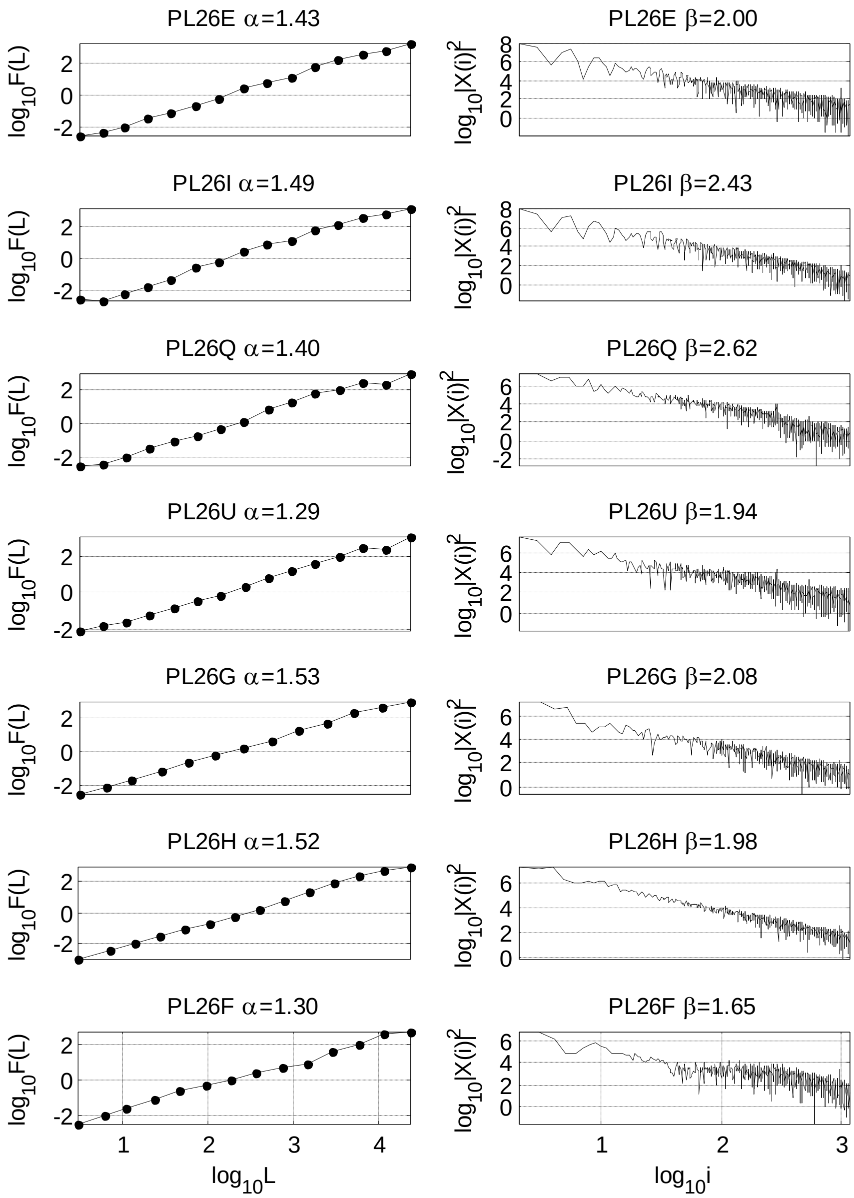

Figure 4: Detrended fluctuation analysis (DFA, α estimate) and power spectral density (PSD, β estimate) of scaling behaviour of the water level time series.

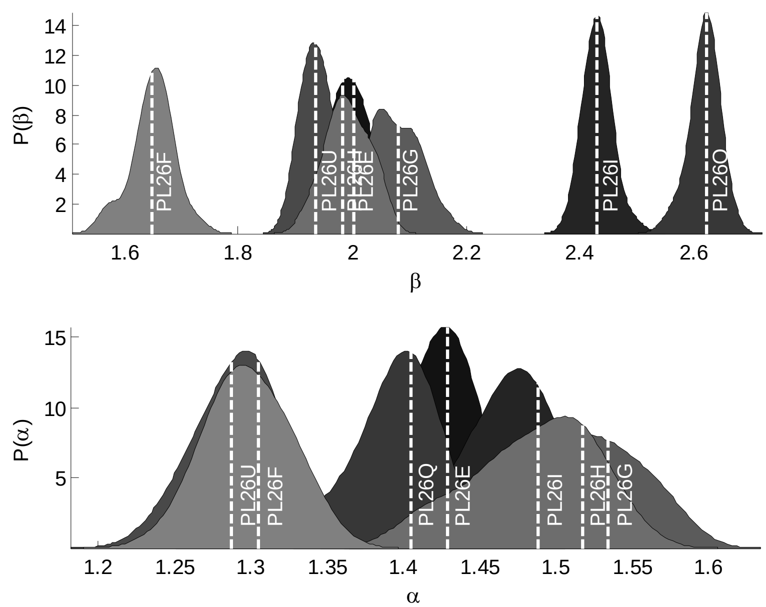

Figure 5: Distribution of scaling exponents for the water level time series, using Gaussian kernel density estimation. The horizontal axis is the exponent, the vertical axis probability. The top panel is the power spectral exponent β, on 20 realizations of PSM noise. Bottom panel is the geometric spectral exponent α, on 20 realizations of RMD noise.