

1 **Robust evidence for random fractal scaling of groundwater levels in**
2 **unconfined aquifers**

3 Max A. Little¹, John P. Bloomfield^{2*}

4 ¹Systems Analysis, Modelling and Prediction Group, University of Oxford, UK

5 ²British Geological Survey, Wallingford, Oxford, UK

6

7 *Corresponding author:

8 Dr John Paul Bloomfield, British Geological Survey, Maclean Building, Crowmarsh Gifford,
9 Wallingford, Oxfordshire, OX10 8BB, UK.

10 Tel. +44 (0)1491 692310 (direct)

11 Tel. +44 (0)1491 838800 (switchboard)

12 Fax. +44(0)1491 692345

13 E-mail jpb@bgs.ac.uk

14

15 (Manuscript submitted to the Journal of Hydrology, Thursday 6th August, 2009)

16

17

18 **Abstract**

19 This study introduces new approaches to improve the statistical robustness of techniques for
20 quantifying the fractal scaling of groundwater levels, and uses these techniques to investigate
21 scaling of groundwater levels from a consolidated permeable carbonate aquifer. Six groundwater
22 level time series and an associated river stage time series from the unconfined Chalk aquifer (a
23 dual-porosity, fractured limestone aquifer) in the Pang-Lambourn catchment, UK, have been
24 analysed. Surrogate data of time series with known scaling properties have been used to estimate
25 the probability distribution of the spectral and geometric scaling exponents determined by
26 Detrended Fluctuation Analysis (DFA) and Power Spectral Density (PSD) respectively; robust
27 regression techniques have been used to improve estimates of the scaling exponents; and robust
28 non-parametric techniques have been used to correlate scaling exponents with features of the
29 boreholes and catchments. Strong statistical support has been found for temporal scaling of
30 groundwater levels over a wide range of time scales, however, bootstrap estimates of the scaling
31 exponents indicate a much larger range of exponents than found by previous studies, suggesting
32 that the uncertainty in existing estimates of scaling exponents may be too small. There is robust
33 evidence that geometrical scaling properties at each borehole can be related to the depth of the
34 observation boreholes and distance of those boreholes from the river in the catchment, but no
35 such correlations were found for the spectral scaling exponents. The results build on the body of
36 evidence that groundwater levels, as with many hydrogeological phenomena, may be well
37 modeled with mathematical concepts from statistical mechanics that do not attempt to capture
38 every detail of these highly heterogeneous and complex systems.

39

40 **KEYWORDS:** groundwater level; fractal scaling; spectral analysis; detrended fluctuation
41 analysis;

42

43 **Introduction**

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

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

71 Such problems with process models are not unique to hydrogeology: in meteorology for
72 example, it has long been recognized that chaos and model error fundamentally limit prediction

73 (Eady, 1951; Lorenz, 1963). The contemporary solution is essentially probabilistic: predictions
74 are made that attempt to represent the full uncertainty due to chaos, produced by many
75 randomized perturbations of the boundary conditions and model equations called *ensemble*
76 *methods* (Buizza 2003). The successes of this approach have precipitated a major conceptual
77 shift from deterministic to probabilistic modelling.

78 This shift may help to mitigate the mathematical limitations of process hydrogeological
79 predictions, but it is not clear that this can also satisfactorily address the effect of high spatial
80 heterogeneity coupled with inaccessible parameters, variables and boundary conditions (Beven,
81 2006). In practice, this may make it impossible to produce detailed predictions of groundwater
82 levels with the same accuracy as, for example, daily surface temperatures. It may well be that the
83 successes of groundwater level predictions resulting from process models calibrated against a
84 few aquifer measurements could be the result of *overfitting*: that is, these predictions are accurate
85 under limited conditions such as short time intervals or locations close to the borehole, but are
86 erroneous for longer intervals or unmeasured sites.

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

96 Classical statistical mechanics explains the bulk statistical properties of simple systems
97 such as ideal gasses. However, many, more complex, systems from diverse disciplinary origins
98 show remarkably similar *scale-invariant* statistical fluctuations of their state variables. These
99 fluctuations are *statistically self-affine* at all length scales, and this is one defining property of
100 *stochastic fractals* (Falconer, 2003). Time series which have stochastic fractal noise, with power
101 spectral density that scales as $f^{-\beta}$, where f is frequency and β is the *spectral scaling exponent*,
102 have been observed from diverse disciplines. This has prompted theoretical explanations such as

103 *self-organised criticality* (SOC) (Bak et al., 1988), *expansion-modification systems* (Li 1991),
104 and *lattice gas density fluctuations* (Jensen, 1990). For example, SOC proposes that under
105 constant small input flux, a local storage mechanism overflows into neighboring regions upon
106 exceeding a capacity threshold. This situation causes cascading overflows on all length scales:
107 time series from these simple models show scaling behaviour which is insensitive to variations in
108 the model parameters. This suggests that this scaling behaviour is in some senses a universal
109 property of complex media.

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

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

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

147 **Methods**

148 Our first task is to assess the evidence for scaling behaviour in borehole data. Firstly, we describe
149 the classical formalism for stochastic fractal time series, which will allow analytical
150 comparisons. We are interested in the class of time series, $x(t)$, that are Gaussian stochastic
151 processes (that is, a set of Gaussian random variables indexed by the real time index t), with the
152 property that $\text{var}[x(t_1) - x(t_2)] \propto |t_1 - t_2|^{2H}$ for arbitrary time indices t_1, t_2 . This condition implies
153 that $x(t)$ and $s^{-H}x(st)$ have the same distribution, for all scale factors $s > 0$. The parameter H is the
154 *scaling exponent* (also known as the *Hurst exponent*) of the self-similar process. As H increases,
155 the resulting stochastic time series becomes smoother. Since the *autocorrelation* can be
156 calculated directly from this definition, it is also straightforward to show (Falconer, 2003) that
157 the power spectrum $X(f) = f^{-\beta}$, where the spectral scaling exponent $\beta = 2H + 1$. The
158 measurements in this study are available at fixed time intervals, where $t_n = n \cdot \Delta t$, i.e. we have x_n
159 $= x(t_n)$. We can simulate approximately self-similar Gaussian time series at these time points
160 t_n using the inverse discrete-time Fourier transform (hereafter, this is referred to as the *power*
161 *spectral method*, PSM). Furthermore, we can estimate the spectral scaling exponent using the
162 forward discrete-time Fourier transform, and since $-\log X(f)/\log f = \beta$, the slope of the log-log

163 plot of f against $X(f)$ is an estimate of the spectral scaling exponent. We call this the power
164 spectral density (PSD) scaling exponent estimation method.

165 The statistical self-similarity of the time series suggests an alternative formalism related
166 to *broken-line processes* (Bras and Rodríguez-Iturbe, 1985). The *random midpoint displacement*
167 (RMD) algorithm can simulate approximately self-affine Gaussian time series on t_n . For a time
168 series of length N that is a power of two, it involves successive subdivision in stages numbered k
169 $= 1, 2, \dots, \log_2 N$, and in the first stage the midpoint is set to $x_{N/2} = 1/2 (x_N + x_1) + \varepsilon$, where ε is a
170 Gaussian random variable of zero mean. We then linearly interpolate the time points between $[1,$
171 $N/2]$ and between $[N/2, N]$. The next stage, $k = 2$, sets the new midpoints $N/4, 3N/4$ according to
172 the same random midpoint displacement scheme. This process repeats until all time points are
173 calculated. The variance of ε at each stage is set to $[1 - 2^{2H-2}] / 2^{2kH}$.

174 Similarly, the scaling exponent of self-affine time series can also be estimated using
175 successive subdivision. By definition, the standard deviation (*fluctuation*) over any sub-interval
176 of length L of the Gaussian time series, will be approximately L^H (Falconer, 2003). Therefore, we
177 can estimate H by first dividing up the time series into sub-intervals of length L , estimating the
178 variance of each sub-interval, and averaging over each standard deviation estimate. Then, by
179 increasing L and repeating the standard deviation calculations over this new sub-interval size, we
180 can estimate H by the slope of the log-log plot of L against the average standard deviation of sub-
181 intervals at each L . *Detrended fluctuation analysis* (DFA) proposes two advances over this basic
182 algorithm. Firstly, although self-affine time series are essentially unbounded (as the variance
183 increases with L), groundwater level time series are bounded, so that estimates of the larger
184 scales are poor. By integrating the time series with the mean removed, i.e. by calculating
185 $\hat{x}_n = \sum_{i=1}^n (x_i - E[x])$, estimates of the scaling at larger sub-intervals are improved. Secondly,
186 most groundwater level time series have trends and other local variations due to factors such as
187 climate variation. By removing local linear trends in each sub-interval (by fitting a straight line
188 or higher-order polynomial to the integrated time series \hat{x}_n and subtracting this), estimates of H
189 insensitive to these trends can be obtained. It can be shown (Heneghan and McDarby, 2000) α ,
190 the spectral scaling exponent, given by the slope of the log-log plot of L against the average
191 standard deviation $F(L)$ is equal to $H - 1$, which is the effect of integrating to obtain \hat{x}_n .

192 Here we introduce an innovation to obtain more robust estimates of groundwater level
193 scaling exponents α and β . Reliable scaling exponent estimates generally require that the log-log
194 plots lie on a straight line (are collinear) over a very large range of length scales. This is often
195 difficult to obtain in practice, because most groundwater level time series are short or have
196 measurement error that may well be temporally correlated. Either the smallest or largest scales
197 will be unusable, or there may be length scales that are *outliers*, in the sense that although most
198 of the points are collinear, a few points are not, and we wish to discard these when using line
199 fitting to estimate the slope. The literature on techniques for addressing this problem of spurious
200 data coming from a very different distribution to the rest is called collectively “*robust statistics*”.
201 Throughout this paper, when we use the term “robust” this is the intended meaning. We use
202 *iteratively reweighted least squares* line-fitting with *Huber penalty function* (Hastie et al., 2001),
203 which concentrates the slope estimate only on those points that are most collinear. It should be
204 noted that this effectively circumvents analysis of *crossovers* – potential changes in scaling
205 properties at different time scales – but gives more reliable estimates of the overall scaling
206 behaviour, which is the main aim of this paper. As a demonstration of the value of robust
207 regression in this application, we generated PSM and RMD time series across 21 values of α in
208 the range [0.5, 2.0]. Using ordinary least squares and robust regression, we computed the average
209 error in the DFA estimate of α , over 10 repetitions for each value of α . The *root-mean square*
210 *error* of the estimate of α with ordinary least squares, from PSM and RMD data was 0.17. Using
211 robust regression, the error was reduced to 0.10 (PSM) and 0.07 (RMD), illustrating the fact that
212 robust regression can lead to marked improvements in scaling exponent estimation.

213 With these methods we can obtain, given a single borehole time series, values for the
214 scaling exponents α and β . However, the statistical mechanical hypothesis holds that the
215 groundwater system is effectively stochastic, which implies that the resulting scaling exponents
216 are random variables. Estimates of exponents from a single time series will simply reflect the
217 statistical variation in that set of measurements, and a true representation of the scaling
218 exponents must be given by the distribution over the exponents. Unfortunately, a concise
219 mathematical description of this distribution is lacking and for the purposes of this study it is
220 reasonable to estimate this distribution by computational means, in particular, by bootstrapping
221 with surrogate data. The surrogate data in this case are time series obtained using the PSM and
222 RMD methods generated using scaling exponents estimated using the PSD and DFA methods on

223 the borehole time series data. Note that we cannot rely on PSM or RMD surrogates alone
224 because there are subtle statistical differences between the time series they generate, differences
225 that arise from the algorithmic details (Bras and Rodríguez-Iturbe, 1985). Assessing the extent to
226 which scaling exponent measurement methods are sensitive to these statistical differences is an
227 important issue that, to our knowledge, has not been addressed in the context of groundwater
228 systems.

229 **Data**

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

240 Groundwater levels have been monitored at six locations at the site and the river level has
241 been monitored using a stilling well for up to five years, Table 1. Water levels at the monitoring
242 locations were measured using pressure transducers and data loggers with a measuring range of
243 $10\text{mH}_2\text{O}$ and a measurement resolution of $0.2\text{cmH}_2\text{O}$. The sampling rate was either hourly or at
244 15-minute intervals. The resulting time series lengths varied from $N = 19,750$ to $N = 49,133$,
245 Table 1. The number of missing measurements was at most 0.1% of the total length of each
246 series, therefore these gaps are ignored in subsequent analysis, since this percentage of missing
247 entries is too small to have a statistically detectable effect on the estimated scaling exponents.
248 Each of the seven water level time series has been normalized to the range $[-1, 1]$ for subsequent
249 analysis, Figure 2.

250 **Results**

251 The performance of the PSD and DFA methods on PSM and RMD bootstrap time series are
252 shown in Figure 3. Minimum/maximum values were assessed by generating 100 fractal time
253 series with the same algorithm. As expected, the PSD method performs almost perfectly on PSM
254 noise, because the method of generating the noise and measuring its scaling exponent are
255 essentially the same. On RMD noise, however, the PSD method performs quite poorly for
256 exponents $\beta < 1$ and $\beta > 2$. The DFA method performs well for PSM time series with exponents
257 $\alpha < 1.2$, but otherwise, it shows a significant deviation away from the true value, although the
258 deviation is not as severe as with PSD on RMD noise. Finally, the DFA method performs very
259 well for RMD noise for $\alpha > 0.8$; for $\alpha < 0.8$, there is a significant deviation away from the true
260 value, but again not as severe as with PSD applied to RMD noise. These findings suggest that,
261 except for the PSD method on RMD noise with high β , although an exact value for the scaling
262 exponent is not always possible, the estimated scaling value always increases with the true value,
263 such that comparisons between estimated values are always indicative of a comparison between
264 the underlying, true values.

265 Having described the techniques, we now apply these to the normalized water level time
266 series. Figure 4 shows log-log plots of DFA sub-interval size L against fluctuation $F(L)$, and
267 frequency bin i against power spectral amplitude $|X(i)|^2$ of the time series x_n . It also shows the
268 scaling exponents α and β for each time series (estimated using robust regression for the log-log
269 line fitting). The DFA sub-intervals ranged on a logarithmic scale from $L = 4$ to $L = N/2$ points.
270 This range is chosen for computational reasons: we need enough points to get a reasonable
271 estimate of the trend fitting (hence at least $L = 4$), and enough DFA sub-intervals for the
272 fluctuation estimates to be reliable (requiring at most $L = N/2$). The PSD frequency bins ranged
273 from $i = 2$ to $i = N/20$, because the power spectral scaling did not extend past this range of
274 frequencies.

275 Figure 4 shows that, over the range of time scales where scaling behaviour could be
276 reliably estimated, the data can be well modeled by a random fractal stochastic process, both in
277 terms of spectral (PSD) and geometric (DFA) scaling. In order to assess whether there is any
278 statistically significant difference between these exponents on the different water level time
279 series, we generated a new set of 20 realizations of stochastic processes with the same scaling

280 exponents as estimated from the data. Since the DFA method is most reliable on RMD noise, we
281 used RMD realizations for α estimates, and for the same reason, for the PSD method used PSM
282 realizations for the β estimates. Figure 5 shows the result, where the distributions are obtained
283 using Gaussian kernel density estimation. This shows that the distributions of the scaling
284 exponents are clearly quite different for each water level time series, for both spectral and
285 geometric exponents. Using the non-parametric Kolmogorov-Smirnov test, we find that all
286 distributions are significantly different ($p < 0.05$, $n = 20$) for all pairwise combinations of water
287 level series. The obtained values of the scaling exponents are summarized in Table 2.

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

294 **Discussion**

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

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

316 Also in agreement with Li and Zhang (2007), we found that the geometrical ‘roughness’
317 of the time series decreases with increasing distance from one of the external driving sources
318 (here, the river flow), which is physically intuitive because the aquifer is a storage medium that
319 tends to ‘dampen’ short-time variations in driving variables. We quantified this relationship as
320 being particularly strong, with a correlation coefficient $r > 0.8$. A novel finding is that this
321 relationship is not detected in spectral scaling exponents. Considering that for some of the
322 boreholes, the estimated spectral scaling exponents $\beta > 2.0$, and given the poor performance of
323 spectral methods on RMD noise for such high values of β , we caution against any physical
324 interpretations of these results based on spectral methods that are sensitive only to statistical
325 means and covariances. Although our findings agree with that of Li and Zhang, the evidence
326 presented here suggests that classical linear spectral analysis cannot reliably extract sufficient
327 information from groundwater time series to detect relationships between scaling behaviour and
328 aspects of the physical configuration of catchments.

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

334 **Conclusions**

335 Our main conclusion is that these results provide a sound statistical basis for supporting the
336 investigation of simple statistical mechanical models as *highly parsimonious* dynamical
337 explanations for the behaviour of groundwater levels. Classical linear models for groundwater
338 flow would be unable to parsimoniously represent fractal scaling – a statistical mechanical model

339 may actually be simpler, because classical linear systems require *infinite memory* to replicate the
340 self-similar behaviour of the measured groundwater levels, whereas nonlinear models require
341 only finite memory. More details of this line of reasoning can be found in Bras and Rodriguez-
342 Iturbe (1985).

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

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

354

355 **Acknowledgements**

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

360 **References**

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

364

365 Bak, P., C. 1996. *How nature works: the science of self-organised criticality*. New York,
366 Springer-Verlag.

367

368 Bak, P., C. Tang and K. Wiesenfeld. 1988. Self-organized criticality. *Physical Review A*, **38**, 364
369 - 373.

370

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

373

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

376

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

379

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

382

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

385

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

388

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

391

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

395

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

399
400 Hamed, K. H. 2007. Improved finite-sample Hurst exponent estimates using rescaled range
401 analysis. *Water Resources Research* **43**, W04413, doi:10.1029/2006WR005111.
402
403 Hastie, T., R. Tibshirani and J. H. Friedman. 2001. *The elements of statistical learning : data*
404 *mining, inference, and prediction : with 200 full-color illustrations*. New York, Springer.
405
406 Heneghen, C. and McDarby, G. 2000. Establishing the relation between detrended fluctuation
407 analysis and power spectral density analysis for stochastic processes. *Phys. Rev. E.* **62**, 6103-
408 6110
409
410 Hurst, H. E. 1951. Long-Term Storage Capacity of Reservoirs. *Transactions of the American*
411 *Society of Civil Engineers*, **116**, 770-799.
412
413 Jensen, H. J. 1990. Lattice Gas as a Model of 1/F Noise. *Physical Review Letters*, **64**, 3103-3106.
414
415 Kantelhardt, J. W., E. Koscielny-Bunde, D. Rybski, P. Braun, A. Bunde and S. Havlin. 2006.
416 Long-term persistence and multifractality of precipitation and river runoff records. *Journal of*
417 *Geophysical Research-Atmospheres*, **111**, D01106, doi:10.1029/2005JD005881.
418
419 Kirchner J.W., X. Feng, and C. Neal. 2001. Catchment-scale advection and dispersion as a
420 mechanism for fractal scaling in stream tracer concentrations. *Journal of Hydrology*, **254**, 82-
421 101.
422
423 Koscielny-Bunde, E., J. W. Kantelhardt, P. Braun, A. Bunde and S. Havlin. 2006. Long-term
424 persistence and multifractality of river runoff records: Detrended fluctuation studies. *Journal of*
425 *Hydrology*, **322**, 120-137.
426
427 Li, W. T. 1991. Expansion-Modification Systems - a Model for Spatial 1/F Spectra. *Physical*
428 *Review A*, **43**, 5240-5260.
429

430 Li, Z. W. and Y. K. Zhang. 2007. Quantifying fractal dynamics of groundwater systems with
431 detrended fluctuation analysis. *Journal of Hydrology*, **336**, 139-146.
432

433 Lorenz, E. N. 1963. Deterministic Nonperiodic Flow. *Journal of the Atmospheric Sciences*, **20**,
434 130-141.
435

436 Lovejoy, S. and D. Schertzer. 1985. Generalized Scale-Invariance in the Atmosphere and Fractal
437 Models of Rain. *Water Resources Research*, **21**, 1233-1250.
438

439 Mandelbrot, B. B. and J. R. Wallis. 1969. Robustness of Rescaled Range R/S in Measurement of
440 Noncyclic Long Run Statistical Dependence. *Water Resources Research*, **5**, 967-988.
441

442 McSharry, P. E. and L. A. Smith. 2004. Consistent nonlinear dynamics: identifying model
443 inadequacy. *Physica D-Nonlinear Phenomena*, **192**, 1-22.
444

445 Murray, B. and M. A. Fonstad. 2007. Preface: Complexity (and simplicity) in landscapes.,
446 *Geomorphology*, **91**, 173-177.
447

448 Olami, Z., H. J. S. Feder and K. Christensen. 1992. Self-organized criticality in a continuous,
449 nonconservative cellular automaton modeling earthquakes. *Physical Review Letters*, **68**, 1244-
450 1247.
451

452 Pelletier, J. D. and D. L. Turcotte. 1997. Long-range persistence in climatological and
453 hydrological time series: analysis, modeling and application to drought hazard assessment.
454 *Journal of Hydrology*, **203**, 198-208.
455

456 Phillips J., D. 2006. Evolutionary geomorphology: thresholds and nonlinearity in landform
457 response to environmental change. *Hydrology and Earth System Science*, **10**, 731-742.
458

459 Ruelle, D. 1984. *Thermodynamic formalism : the mathematical structures of classical*
460 *equilibrium statistical mechanics*. New York, NY, USA, Cambridge University Press.

461
462 Smale, S. 1967. Differentiable Dynamical Systems. I. Diffeomorphisms. *Bulletin of the*
463 *American Mathematical Society*, **73**, 747-817.
464
465 Tessier, Y., S. Lovejoy, P. Hubert, D. Schertzer and S. Pecknold. 1996. Multifractal analysis and
466 modeling of rainfall and river flows and scaling, causal transfer functions. *Journal of*
467 *Geophysical Research-Atmospheres*, **101**, 26427-26440.
468
469 Wells, J. T., D. R. Janecky and B. J. Travis. 1991. A Lattice Gas Automata Model for
470 Heterogeneous Chemical-Reactions at Mineral Surfaces and in Pore Networks. *Physica D* **47**,
471 115-123.
472
473 Zhang, Y. K. and Z. W. Li. 2005. Temporal scaling of hydraulic head fluctuations:
474 Nonstationary spectral analyses and numerical simulations. *Water Resources Research*, **41**,
475 W07031 , 10.1029/2004WR003797.
476
477 Zhang, Y. K. and Z. W. Li. 2006. Effect of temporally correlated recharge on fluctuations of
478 groundwater levels. *Water Resources Research*, **42**, W10412, 10.1029/2005WR004828.
479
480 Zhang, Y. K. and K. Schilling. 2004. Temporal scaling of hydraulic head and river base flow and
481 its implication for groundwater recharge. *Water Resources Research*, **40**, W03504,
482 10.1029/2003WR002094.
483
484

485 **Tables & Figures**

486

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

488

489 Table 2. Median stochastic fractal scaling exponents α , β obtained for water level time series
490 from the Boxford site. The confidence intervals are inter-quartile range, estimated using 20
491 realizations of stochastic processes with the same scaling exponents as estimated from the data.

492

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

496

497 Figure 1. Schematic illustration of the Boxford site showing the relative locations of the
498 boreholes and the stilling well in the river Lambourn (PL26U).

499

500 Figure 2. Normalised water level (NWL) time series from the Boxford site. The vertical axis is
501 unitless, the horizontal axis is time in days since the start of the record, excluding missing
502 measurements.

503

504 Figure 3. Performance of power-spectral density (estimating β) and detrended fluctuation
505 analysis (estimating α) methods on power-spectral (PSM) and random midpoint (RMD) noise.

506

507 Figure 4. Periodograms of the results of the detrended fluctuation analysis (DFA, α estimate) and
508 power spectral density (PSD, β estimate) estimates showing scaling behaviour of the water level
509 time series.

510

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

<i>Borehole ID (measurement type)</i>	<i>Easting</i>	<i>Northing</i>	<i>Distance from PL26U (m)</i>	<i>Distance from river (m)</i>	<i>Observation zone depth (m bGL)</i>	<i>Geology</i>	<i>Land cover</i>	<i>Sample rate (mins)</i>	<i>Record start</i>	<i>Record end</i>	<i>Missing data</i>
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	

516 **Table 1.**

517

518

519 **Table 2.**

520

521

522

	<i>PL26E</i>	<i>PL26I</i>	<i>PL26Q</i>	<i>PL26U</i>	<i>PL26G</i>	<i>PL26H</i>	<i>PL26F</i>
β	2.01±0.03	2.43±0.03	2.62±0.03	1.94±0.03	2.08±0.04	1.98±0.04	1.65±0.06
α	1.42±0.02	1.48±0.05	1.40±0.03	1.29±0.02	1.49±0.05	1.52±0.04	1.30±0.02

523 **Table 3.**

524

525

	<i>Distance from PL26U against α</i>	<i>Distance from river against α</i>	<i>Depth of observation zone against α</i>	<i>Distance from PL26U against β</i>	<i>Distance from river against β</i>	<i>Depth of observation zone against β</i>
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

526
527
528
529
530
531
532
533

534

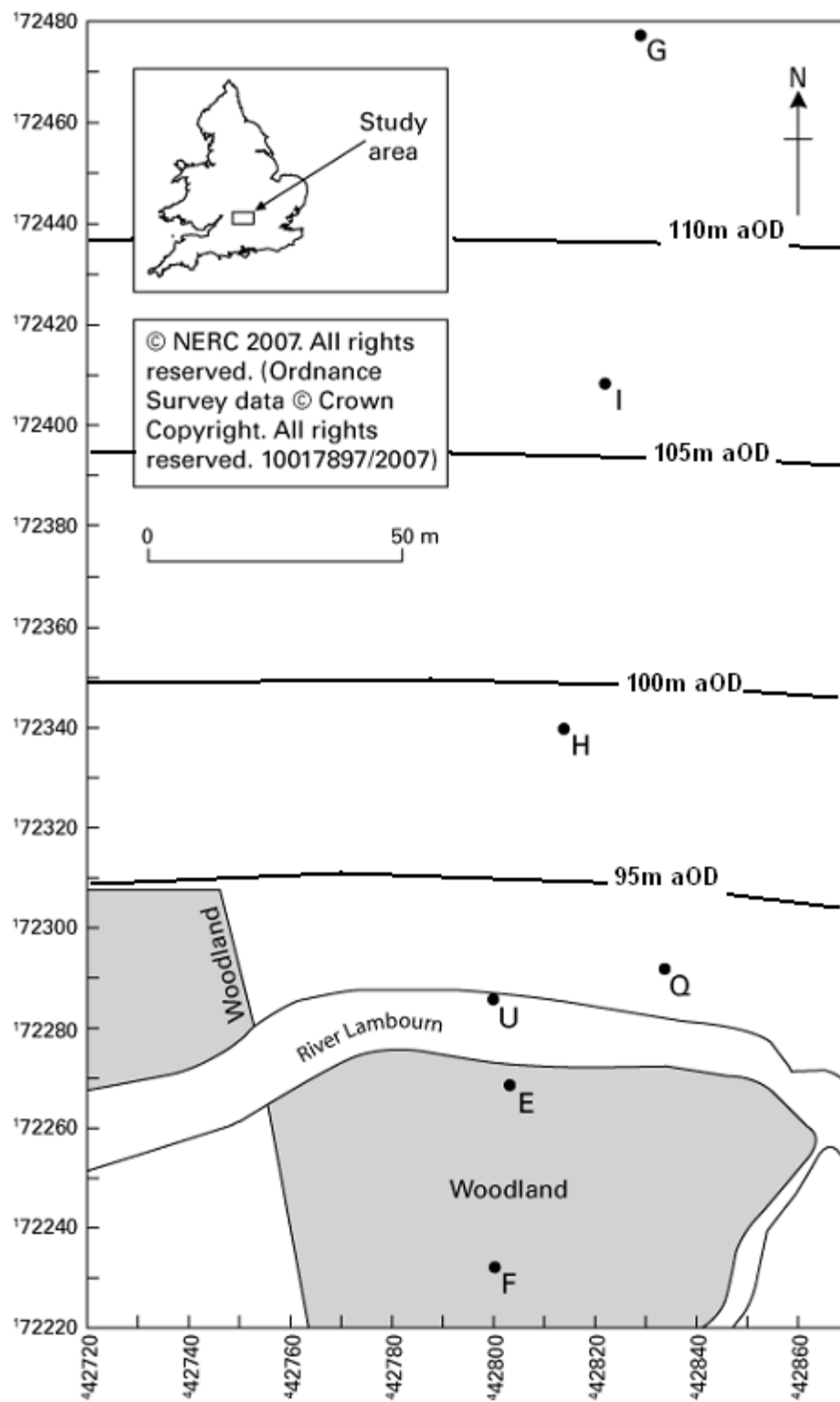
535

536

537

538

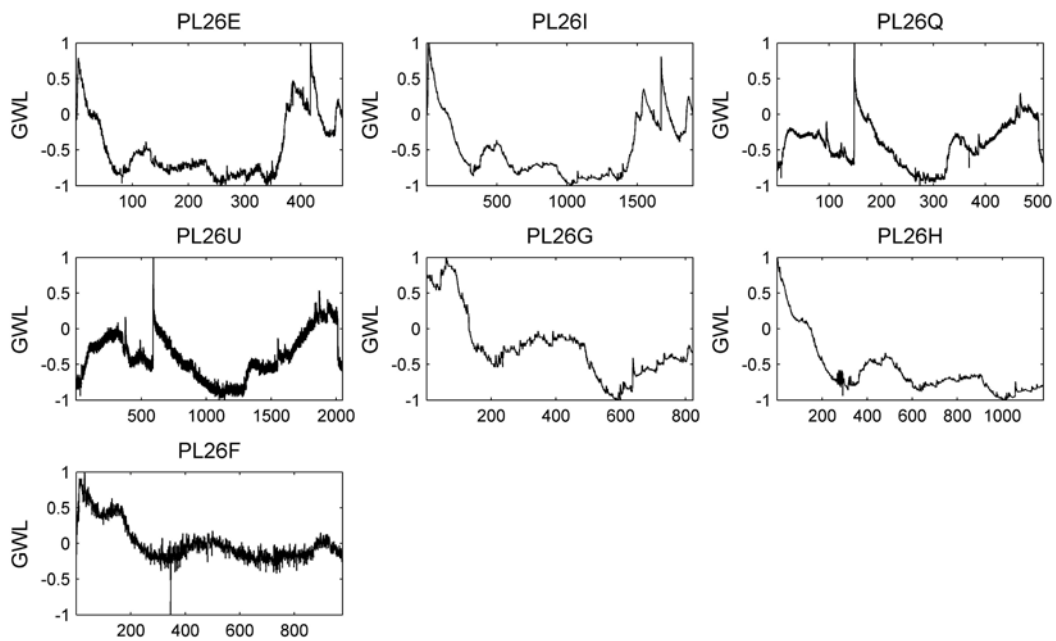
539 Figure 1.



540
541
542

543 Figure 2.

544



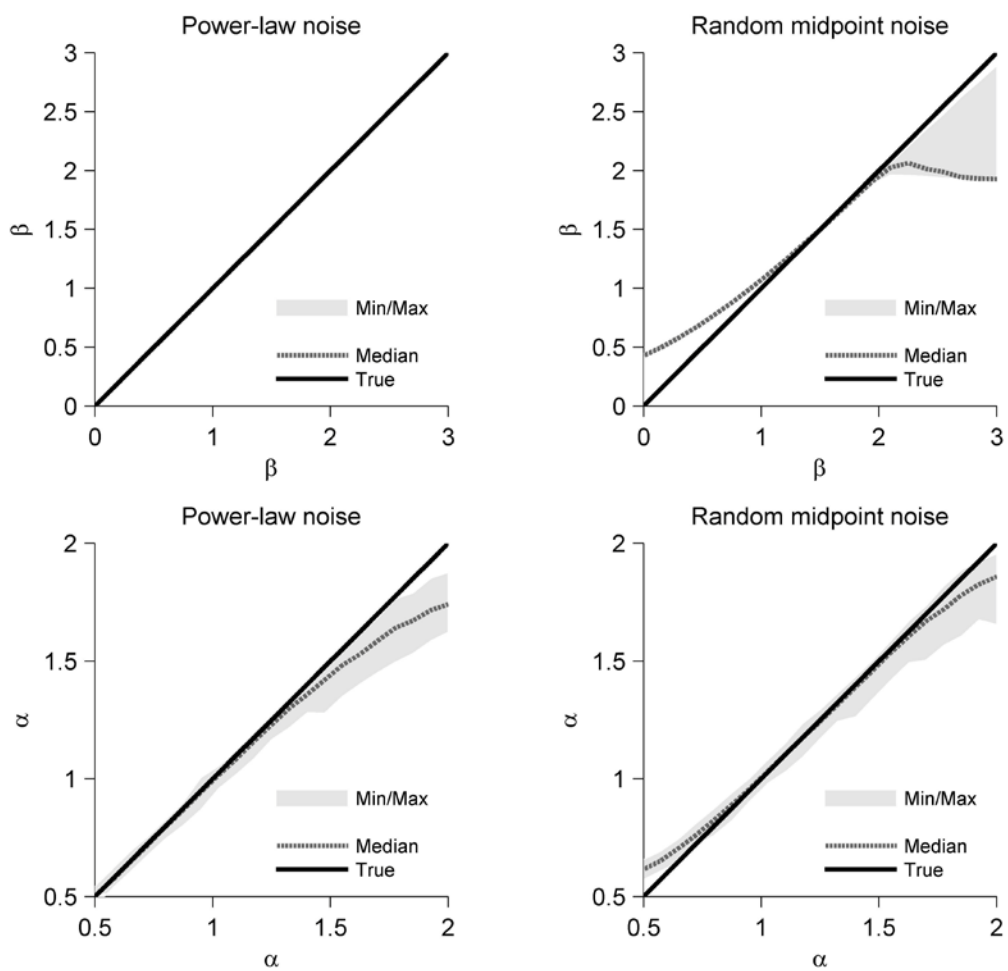
545

546

547

548 Figure 3.

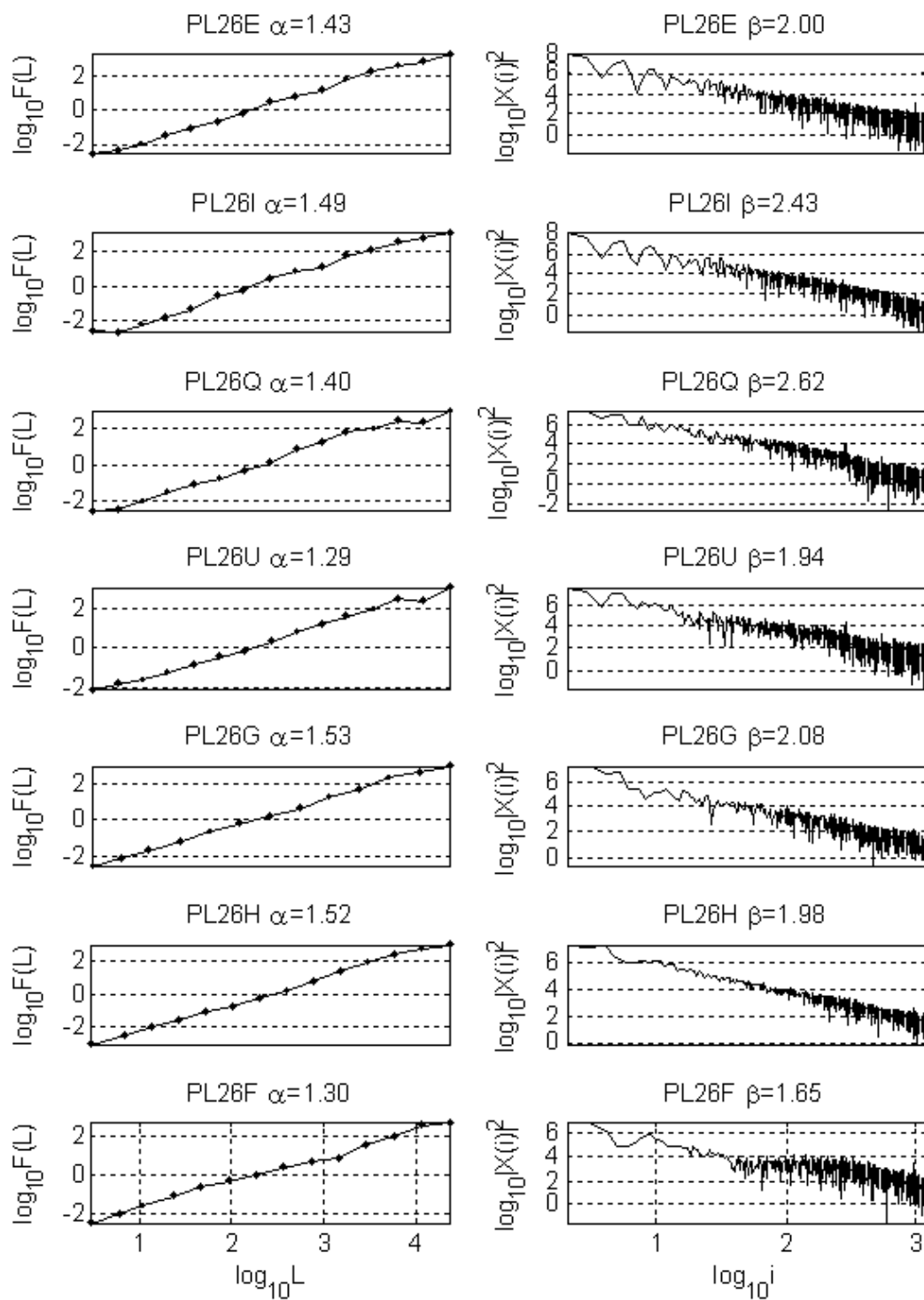
549



550

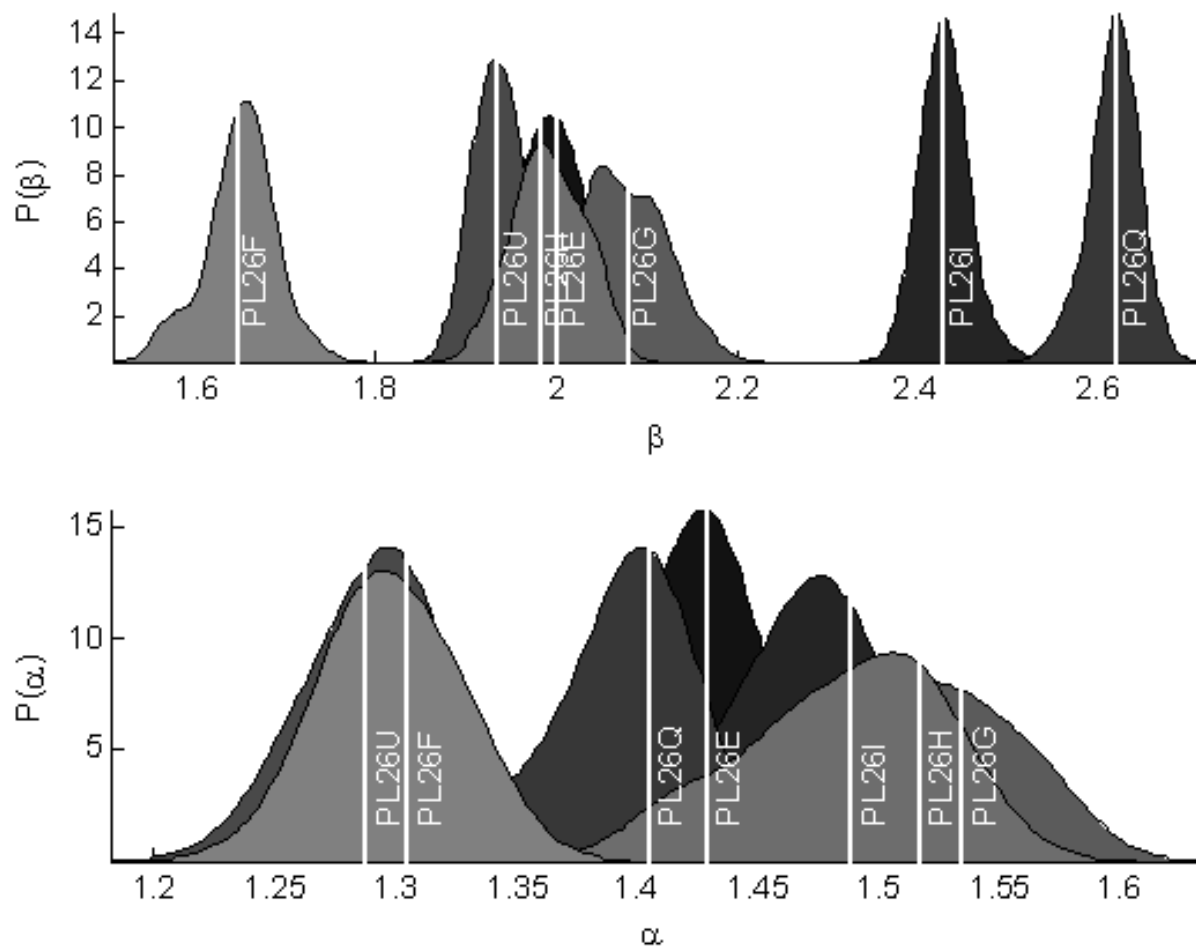
551

552



557 Figure 5.

558



559