Multifractal Modelling of Porosity in Heterogeneous Aquifers

V. V. Anh[1], N. Arunakumaren[2], K. Bajracharya[2], Q. Tieng[1]and I. Turner[1]

[1]School of Mathematical Sciences
Queensland University of Technology
GPO Box 2434, Brisbane, Q. 4001, Australia

[2]Queensland Department of Natural Resources,
Block C, Level 2, Resource Sciences Centre,
80 Meiers Road, Indooroopilly, Q. 4068, Australia

Abstract This paper presents a method to estimate the long-range dependence and multifractality in the data on porosity observed at two locations of the Bundaberg saltwater monitoring network. The resulting models will be used in the classification and clustering of the network.

1. INTRODUCTION

Over-exploitation of coastal aquifers in Queensland results in saltwater intrusion into the aquifers. The Queensland Department of Natural Resources is addressing the problem in the Bundaberg area, in part, by substituting surface water for groundwater in problem areas and by applying severe water restrictions. There is a need for developing a reliable model for prediction of saltwater intrusion which enhances resource management decisions and maintains production from important aquifers.

Management of many aquifer systems takes place on a number of scales ranging from the regional scale of tens of kilometers down to the farm property scale of 0.5-1 kilometer. Hydraulic data from monitoring bores are used to estimate large scale variations of hydraulic conductivity. On the other hand, solute transport models require information on hydraulic conductivity at a much finer scale than that of hydraulic head data. It is therefore essential to be able to fuse hydraulic head data, solute concentrations and geological data to result in a transport model that can represent both fine and large scale aquifer heterogeneities.

This paper will describe the first step in the development of such a model, namely, the characterisation of monitoring data. The current saltwater monitoring network of 260 boreholes covers a large coastal area with various degree of heterogeneity. The characteristics of these monitoring data will be used to identify homogeneous clusters of the region.

We shall rely on the possible long-range dependence (LRD) and multifractality of measurements on porosity and hydraulic conductivity to classify the network. We estimate the LRD exponent in each data series based on the fractional Riesz-Bessel motion model. Multifractal data are modelled as multiplicative cascades generated from a binomial probability distribution on a Cantor set. The multifractality of the data is then inferred via the multifractal formalism. This paper will present some numerical results based on measurements of porosity at two locations of the network. It is the purpose of the paper to demonstrate the potential of this method in clustering the network.

Section 2 will describe the models used in the estimation of the LRD exponent and the multifractality of the data series. Section 3 will present some numerical results on porosity data, and some conclusions will be drawn in Section 4.

2. METHODOLOGY

Let $\{X(t), t \in \mathbb{R}\}$ be a stationary process with mean zero. We shall assume that the spectral density of $X(t)$ has the form

$$f(\lambda) = \frac{c}{|\lambda|^{2\gamma}\left(1+\lambda^2\right)^{\alpha}}, \quad (1)$$

$$c > 0,\ 0 < \gamma < \frac{1}{2},\ \alpha + \gamma > \frac{1}{2},\ \lambda \in \mathbb{R}.$$

It is noted that $f(\lambda)$ has an integrable pole at $\lambda = 0$. Hence, by definition, the process $X(t)$ possesses LRD, and the exponent γ denotes the extent of this LRD. Also, $\int_{\mathbb{R}} f(\lambda)\, d\lambda < \infty$ for $0 < \gamma < 1/2$ and $\alpha + \gamma > 1/2$. Thus, the component $\left(1+\lambda^2\right)^{-\alpha}$ plays an important role in the definition of a spectral density of the form (1).

In fact, $(1+\lambda^2)^{-\alpha}$ is the Fourier transform of the Bessel potential, while $|\lambda|^{-2\gamma}$ is the Fourier transform of the Riesz potential. A process with spectral density of the form (1) is shown to exist and named the Riesz-Bessel motion in Anh et al. [1999a]. It is shown in Anh et al. [1999b] that the exponent α indicates the second-order intermittency of the process $X(t)$.

We shall estimate the parameters $\theta = (c, \gamma, \alpha) \in \Theta = (0,\infty) \times (0, 1/2) \times (0,1)$ of (1) by computing the minimum constrast estimator θ_N defined by

$$\theta_N = \arg\min_{\theta \in \Theta_0} L_N(\theta), \tag{2}$$

where Θ_0 is a compact subset of Θ,

$$L_N(\theta) = \frac{1}{4\pi} \int_{\mathbb{R}} \left(\ln f(\lambda) + \frac{I_N(\lambda)}{f(\lambda)} \right) \frac{d\lambda}{1+\lambda^2}. \tag{3}$$

The objective function (3) is the continuous version of the Gauss-Whittle contrast function. The form (3) is suggested from the entropy formula for a continuous-time process as given in Dym and McKean [1976]. The periodogram $I_N(\lambda)$ in (3) is computed as

$$I_N(\lambda) = \frac{1}{2\pi N} \left| \int_0^N e^{-i\lambda t} X(t)\, dt \right|^2, \tag{4}$$

where $N > 0$ is the upper bound of the interval $[0, N]$ on which $X(t)$ is observed. As established in Gao et al. [1999], as $N \to \infty$,

$$EI_N(\lambda) \to f(\lambda, \theta_0),$$

$$E(I_N(\lambda) - EI_N(\lambda))^2 = f^2(\lambda, \theta_0) + o(1),$$

$$E(I_N(\lambda) - EI_N(\lambda))(I_N(\omega) - EI_N(\omega)) = o(1) \quad \forall \lambda \neq 0,\ \lambda \neq \omega,$$

$$\theta_N \to \theta_0 \text{ with probability 1},$$

where θ_0 is the true value of θ. Thus the method produces a strongly consistent estimator for both LRD and intermittency parameter.

Let us now look at a model for the multifractality of the process $X(t)$. Define $Y(t) = |X(t)| / E|X(t)|$. Then $Y(t) \geq 0$ and $EY(t) = 1$ for all t. For each t in $\mathbb{R}$, let

$$\alpha(t) = \lim_{r \to 0^+} \frac{\log Y(B(t;r))}{\log r}$$

be the local dimension of Y at t, where $B(t;r) = \{s; |s-t| < r\}$ and $Y(B(t;r)) = \int_{B(t;r)} Y(s)\, ds$. Let $G(\alpha) = \{t; \alpha(t) = \alpha\}$ and let $f(\alpha)$ be the Hausdorff dimension of $G(\alpha)$. We call $f(\alpha)$ the *singularity spectrum* of Y, and we say that Y is a *multifractal measure* if $f(\alpha) \neq 0$ for a continuum of α. In this definition, monofractals therefore consist of singularities all of the same strength (i.e. $\alpha(q) = \alpha$, a constant, for all q, and the graph of $f(\alpha)$ consists of one point). On the other hand, multifractals will display heterogeneous scaling, which can be characterised by

$$\sum (Y(t;r))^q \sim r^{\tau(q)},\ q \in \mathbb{R}, \tag{5}$$

as $r \to 0$, where the sum is taken over all disjoint intervals of length r in the smoothing

$$Y(t;r) = \frac{1}{r} \int_{t-\frac{r}{2}}^{t+\frac{r}{2}} Y(s)\, ds,\ r > 0.$$

There is a relationship between $f(\alpha)$ and $\tau(q)$. In fact, let τ^* denote the concave conjugate of τ (also known as the Legendre transform of τ), i.e.,

$$\tau^*(\alpha) = \inf_{q \in \mathbb{R}} \{q\alpha - \tau(q)\}.$$

Hentschel and Procaccia [1983], Halsey et al. [1986] showed heuristically that, if the meaure Y is constructed from a cascade algorithm and if $\tau(q)$ and $f(\alpha)$ are smooth and concave, then $\tau^*(\alpha) = f(\alpha)$ and dually $f^*(q) = \tau(q)$.

This relationship is called the *multifractal formalism*. The multifractal formalism is a useful tool in applications. In fact, it yields that

$$\tau(q) = \sup_{0 < \alpha < \infty} \{f(\alpha) - q\alpha\}.$$

Suppose that the supremum is attained at $\alpha = \alpha(q) > 0$. Then

$$\frac{d}{d\alpha}(f(\alpha) - q\alpha) = 0,$$

which implies

$$q = \frac{df}{d\alpha}(\alpha(q)),$$

and

$$\begin{aligned} \frac{d\tau}{dq} &= \frac{df}{d\alpha}\frac{d\alpha}{dq} - \alpha(q) - q\frac{d\alpha}{dq} \\ &= -\alpha(q). \end{aligned} \tag{6}$$

Given a sample of $X(t)$, the mass exponent $\tau(q)$ can be computed from (5) via a log regression. The function $\alpha(q)$ is then given by (6) and the singularity spectrum $f(\alpha)$ is obtained from

$$\tau(q) = f(\alpha(q)) - q\alpha(q). \tag{7}$$

Our basic assumption is that $Y(t)$ is generated by a multiplicative cascade with a binomial generator characterised by a probability $p, 0 < p \leq \frac{1}{2}$. In other words, according to Falconer [1990],

$$\log \sum (Y(t;r))^q = \log \left(r^{-\log_2(p^q + (1-p)^q)} \right).$$

From (5) and (7) we get

$$\begin{aligned}\tau(q) &= \lim_{r\to 0}\frac{\log\sum(Y(t;r))^q}{\log r}\\ &= -\log_2(p^q+(1-p)^q). \qquad (8)\end{aligned}$$

Define

$$K(q) = -\tau(q) + q - 1. \qquad (9)$$

Then, for the binomial cascade described above,

$$K(q) = \log_2(p^q+(1-p)^q) + q - 1. \qquad (10)$$

Remark 1 *It follows directly from (9) that* $K(0) = 0$, *and since* $EY(t;r) = 1$ *by definition, we also have* $K(1) = 0$. *Writing* Y *for* $Y(t;r)$ *and considering* r *sufficiently small, we get from (9) and (10) that*

$$K(q) = \frac{\log E(Y^q)}{\log r}. \qquad (11)$$

Thus

$$K'(q)\log r = \frac{(E(Y^q\log Y))\log Y}{E(Y^q)},$$

$$\begin{aligned}K''(q)\log r &= (E(Y^q))^2\\ &\times\left(E(Y^q)\left(E\left(Y^q(\log Y)^2\right)\right)\log Y\right.\\ &\left.-(E(Y^q\log Y))^2\log Y\right).\end{aligned}$$

But

$$\begin{aligned}(E(Y^q\log Y))^2 &= \left(E\left(Y^{\frac{q}{2}}Y^{\frac{q}{2}}\log Y\right)\right)^2\\ &\le E(Y^q)\,E\left(Y^q(\log Y)^2\right)\end{aligned}$$

by Schwarz's inequality. Consequently, $K''(q) \ge 0$, *so that* $K(q)$ *is a convex function. It also follows from (11) that* $K(q) < 0$ *iff* $E(Y^q) < 1$, *which holds only if* $0 < q < 1$. *These results are useful when the model (10) is fitted to empirical data.*

Remark 2 *In order to examine directly whether* $Y(t)$ *is monofractal or multifractal, it is more convenient to consider another scaling exponent given by*

$$E|Y(t) - Y(t-r)|^q \sim r^{\zeta(q)},\ q \ge 0,$$

as $r \to 0$. *It is seen that* $\zeta(0) = 0$, *and no other exponent is known* a priori *(in contrast to* $K(q)$, *where there are two* a priori *exponents:* $K(0) = 0$ *and* $K(1) = 0$). *Following the same argument as for* $K(q)$, *it can be shown that* $\zeta''(q) \le 0$; *hence* $\zeta(q)$ *is concave. Furthermore, under the condition that* $Y(t)$ *is bounded, the function* $\zeta(q)$ *is monotonically nondecreasing (Marshak et al. [1994]). These results imply that, if* $Y(t)$ *is a monofractal, its scaling will be simply given by* $\zeta(q) = q\alpha$, *where* α *is a constant. In particular, for fractional Brownian motion with Hurst index* H, *we have* $\zeta(q) = qH$. *This result is a convenient tool to test if* $Y(t)$ *is a monofractal.*

3. EXPERIMENTAL RESULTS

Figure 1 displays the locations of the bores at Gooburrum, which form part of the Bundaberg saltwater monitoring network. In particular, the bores 13700202 and 13500148 are marked on the map. This paper examines the measurements of porosity at these two locations. The measurements are collected at regular intervals of 2cm and going down to 70m deep. The series at bores 13700202 and 13500148 are plotted in Figures 2 and 3 respectively. We assume that they are sample paths of a stationary stochastic process with spectral density of the form (1). The periodogram of the series of Figure 2, computed as in (4), is displayed in Figure 4 in the log-log form (i.e. $\ln f(\lambda)$ against $\ln\lambda$). It shows a negative slope in the low frequency range and a steeper negative slope in the high frequency range; hence the form (1) seems suitable in modelling this data series. The parameters c, γ and α are next estimated by solving problem (2). The following values are obtained: $c = 10.9$, $\gamma = 0.389$, $\alpha = 0.463$. The corresponding estimates for the data series of Figure 3 are $c = 11.6$, $\gamma = 0.348$, $\alpha = 0.506$.

The estimated value of γ means that LRD is present, while the estimated value of α indicates a possible multiple scaling of the data series (instead of a monoscaling behaviour as in self-similar processes). The condition $\alpha+\gamma > 1/2$ is also satisfied, confirming the appropriateness of model (1). The fitted model is shown as the solid line in Figure 4.

We next study the multifractality of the data series. We first compute the exponent $\zeta(q)$ as a function of q as defined in Remark 2 using log regression. The $\zeta(q)$ function is plotted in Figure 5 for the data at bore 13700202 as an example. It can be seen that $\zeta(q)$ is a nonlinear function of q, indicating that the series is multifractal. The exponent $K(q)$ is next computed as in (9) and (5), again using log regression. This is shown in Figure 6 for the data at bore 13700202 and in Figure 7 for the data at bore 13500148. It is seen that the curves display the theoretical convex shape as demonstrated in Remark 1, with zeroes at $q = 0$ and $q = 1$. The model (10) is then fitted using nonlinear least squares to the $K(q)$ curves. The estimates for the probability p generating the multiplicative cascade are $p = 0.4641$ for bore 13700202 and $p = 0.4657$ for bore 13500148. The smaller value of p for bore 13700202 means

that this location is more intermittent than bore 13500148. The different values of the probability p therefore indicate that these two locations have different scaling behaviours.

4. CONCLUSIONS

The experimental results confirm the simultaneous presence of long-range dependence and multifractality in the porosity data. The excellent fit of models (1) and (10) indicates that, even though simple, they are quite adequate in capturing these two important phenomena and distinguishing the different behaviours of the data. The models will be used to characterise the monitoring data at each location of the network. The information distance between any two locations will then be computed using the spectral density (1) and the probability p inferred from model (10). These information distances play a key role in the clustering of the monitoring network. This will be the next stage of the program outlined in the Introduction.

REFERENCES

Anh, V. V., Angulo, J. M. and Ruiz-Medina, M. D., Possible long-range dependence in fractional random fields, *J. Statist. Plann. & Inference*, 80, 95-110, 1999a.

Anh, V. V., Heyde, C. C. and Tieng, Q., Stochastic models for fractal processes, *J. Statist. Plann. & Inference*, 80, 123-135, 1999b.

Dym and McKean, Gaussian Process, Function Theory and the Inverse Spectral Problem, Academic Press, New York,1976.

Falconer, Fractal Geometry, Wiley, 1990.

Gao, G. T., Anh, V. V. and Heyde, C. C., Statistical estimation of fractional Riesz-Bessel motion, (submitted), 1999.

Halsey, T. C., Jensen, M. H., Kadanoff, L. P., Procaccia, I. and Shraiman, B. J., Fractal measures and their singularities: The characterization of strange sets, *Phys. Rev. A*, 33, 1141-1151, 1986.

Hentschel, H. G. E. and Procaccia, I., The infinite number of generalized dimensions of fractals and strange attractors. *Physica D*, 8, 435-444, 1983.

Marshak, A., Davis, A., Cahalan, R. and Wiscombe, W., Bounded cascade models as nonstationary multifractals, *Phys. Rev. E*, 49, 55-69, 1994.

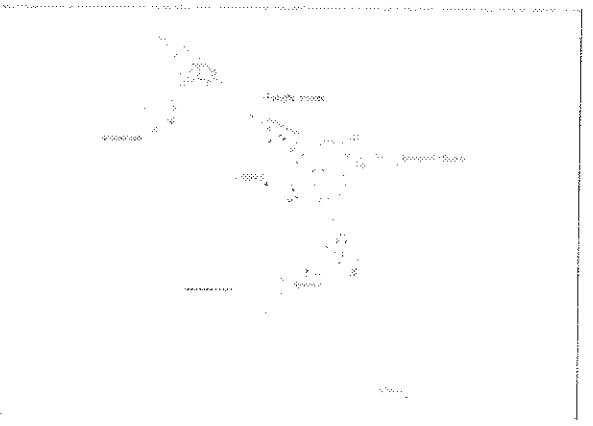

Figure 1: Saltwater monitoring bores at Gooburrum, Bundaberg, with the locations of bores 13500148 and 13700202 indicated.

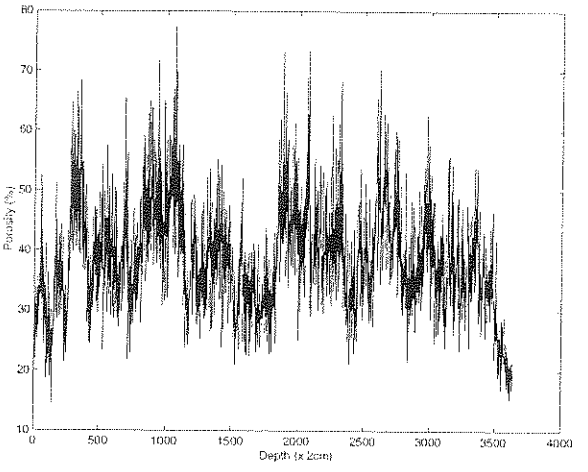

Figure 2: Porosity at bore 13700202 in the Bundaberg irrigation area.

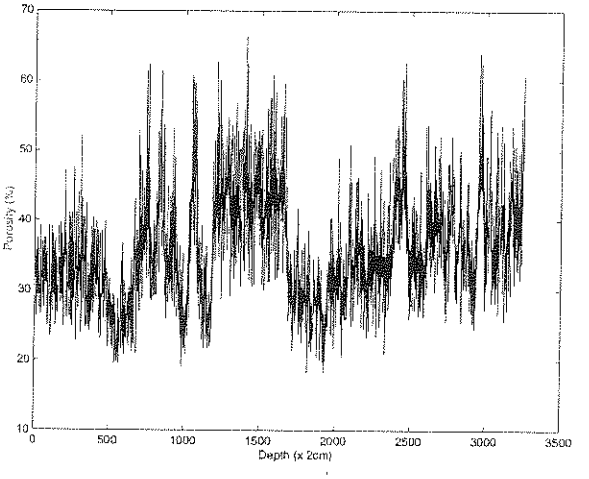

Figure 3: Porosity at bore 13500148 in the Bundaberg irrigation area.

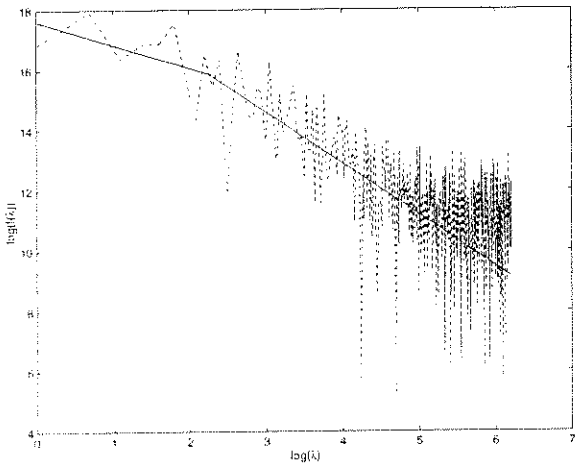

Figure 4: The periodogram and the fitted model (solid line) of the data series of Figure 2.

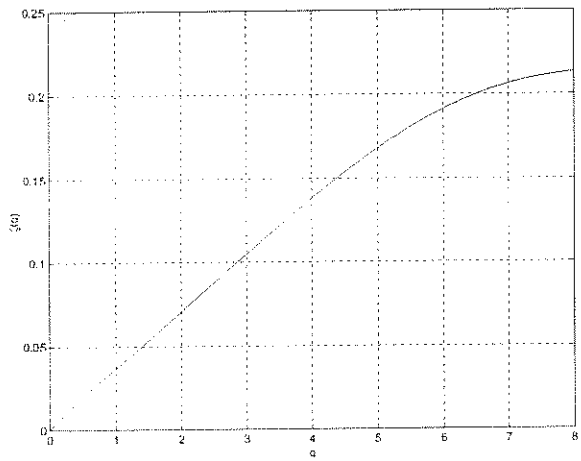

Figure 5: The exponent $\zeta(q)$ of the data series of Figure 2.

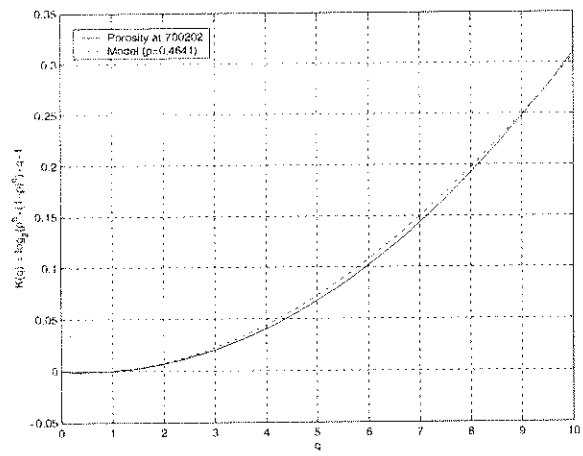

Figure 6: The multifractal exponent $K(q)$ and the fitted model of the data series of Figure 2.

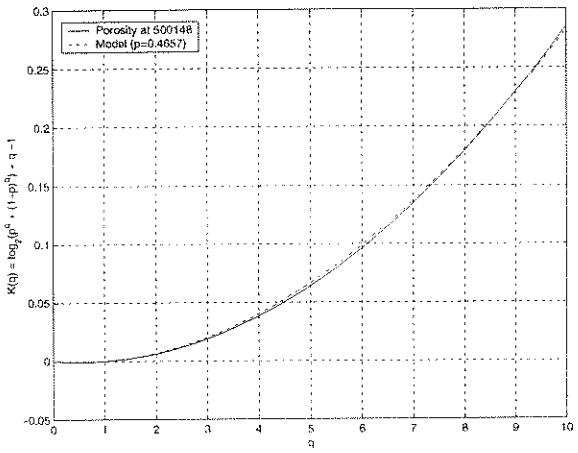

Figure 7: The multifractal exponent $K(q)$ and the fitted model of the data series of Figure 3.