Confidence interval estimation of hydraulic heads at unobserved locations using stationary stochastic models and geologic interpretations

Andreas Langousis, Vagia Xevgeni and Vassilios Kaleris

Department of Civil Engineering, University of Patras, Greece (andlag@alum.mit.edu)

1. Introduction

Effective design of groundwater observation networks, requires accurate and robust hydraulic head estimates at multiple locations in an aquifer. In most groundwater applications, however, the spatial extent of geologic formations is generally assessed based on subjective geologic interpretations in the light of geologic maps. Hence, parameterizing the uncertainty in hydraulic head estimation constitutes an important topic for engineering applications.

In this work, we study how the statistical structure of hydraulic conductivity fields affects the distribution of the absolute error in hydraulic head estimates at unobserved locations in an aquifer. The latter is calculated as a function of: a) the characteristic scale l_c of geologic formations, b) the standardized distance from the nearest measuring locations (i.e. inter-borehole distance), and c) the small scale variability inside each formation.

2. Error distribution in hydraulic head estimation

A. Dimensional analysis for 1D flow in confined aquifers

Suppose a one dimensional (1D) confined aquifer of total length L, formed by r_{max} successive hydraulic conductivity units of equal length $l_0 = L/r_{max}$; see **Figure 1** below.

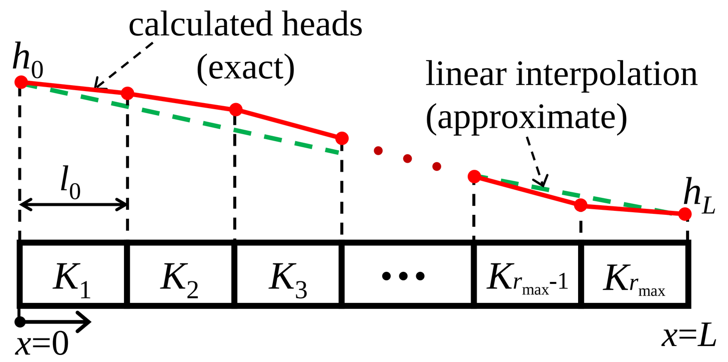

Figure 1: Schematic representation of the calculated (exact; red broken line) and linearly interpolated (green line) hydraulic heads in a 1D confined aquifer, formed by r_{max} successive hydraulic conductivity units of equal length.

In the case when the hydraulic conductivities K_i, i = 1, 2, …, r_{max} are known, one can calculate the exact hydraulic head $h(x)$ at any location x in the direction of the flow (see red broken line in **Figure 1**), as:

$$h(x) = h_0 - q l_0 \left[\sum_{i=1}^{s} \frac{1}{K_i} + \frac{(x/l_0 - s)}{K_{s+1}} \right], \quad x \in [0, L]$$

where s = int(x/l_0) is the integer part of the ratio x/l_0, $h_0 = h(x = 0)$, $h_L = h(x = L)$, and:

$$q = \left(\sum_{i=1}^{r_{max}} \frac{1}{K_i} \right)^{-1} \frac{h_0 - h_L}{l_0}$$

is the groundwater discharge per unit width of the aquifer (i.e. perpendicular to the direction of the flow).

In the lack of hydraulic conductivity information, one can obtain an estimate $\hat{h}(x)$ of the standardized hydraulic head $h(x)$ by linearly interpolating between the two measuring locations:

$$\hat{h}(x) = h_0 - \frac{x}{L}(h_0 - h_L), \quad x \in [0, L]$$

Under this setting, one may define the **standardized absolute error $|e(u)|$** as:

$$|e(u)| = \left| \frac{h(u) - \hat{h}(u)}{h_0 - h_L} \right|, \quad u = x/L \in [0, 1]$$

B. Theoretical attributes of the distribution of $|e(u)|$

Independent of the geologic structure of the aquifer (see **Figure 1** above), the standardized absolute error $|e(u)|$ exhibits statistical symmetry; see **Figure 2**.

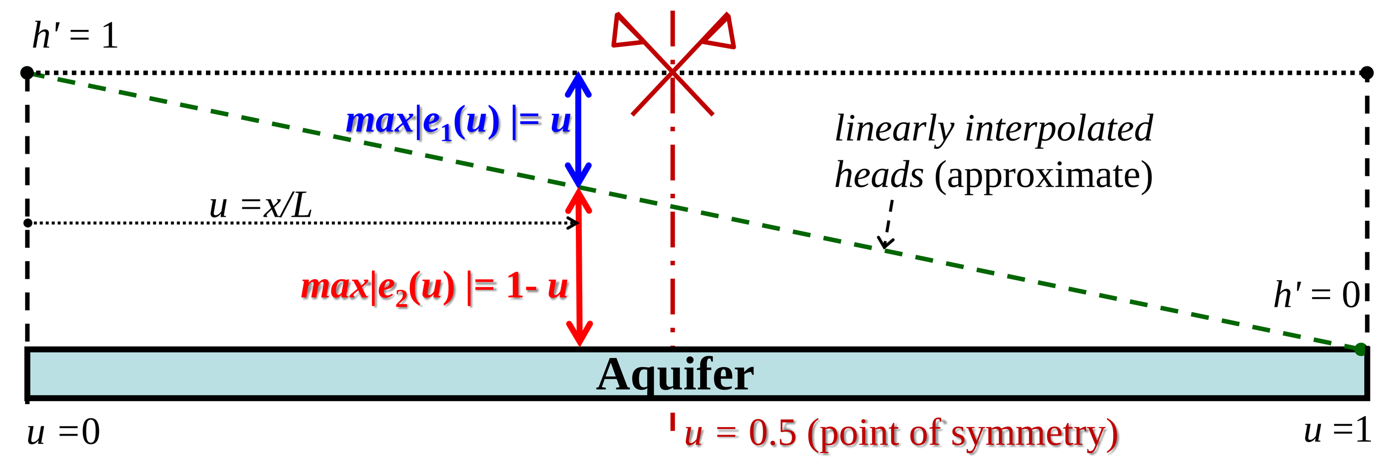

Figure 2: Schematic illustration of statistical symmetry in the study problem. $h'(u) = (h(u) - h_L)/(h_0 - h_L)$ is the standardized hydraulic head.

It follows from statistical symmetry (see e.g. **Figure 2**), and simple geometric interpretations, that:

- The cumulative distribution function (CDF) of the standardized absolute error $|e(u)|$ satisfies (see **Figure 2** above):

$$F_{|e(u)|} = F_{|e(1-u)|}, \quad u = x/L \in [0, ½]$$

- Due to the **geometry** of the problem under consideration, for any location $u = x/L \in [0, ½]$ along the aquifer, $|e(u)|$ is described by a two component distribution (see **Figure 3**):

$$\left.\begin{array}{ll} \textbf{Component 1:} & 0 \le |e_1(u)| \le u \\ \textbf{Component 2:} & u < |e_2(u)| \le 1 - u \end{array}\right\} \text{ for any } u = x/L \in [0, ½]$$

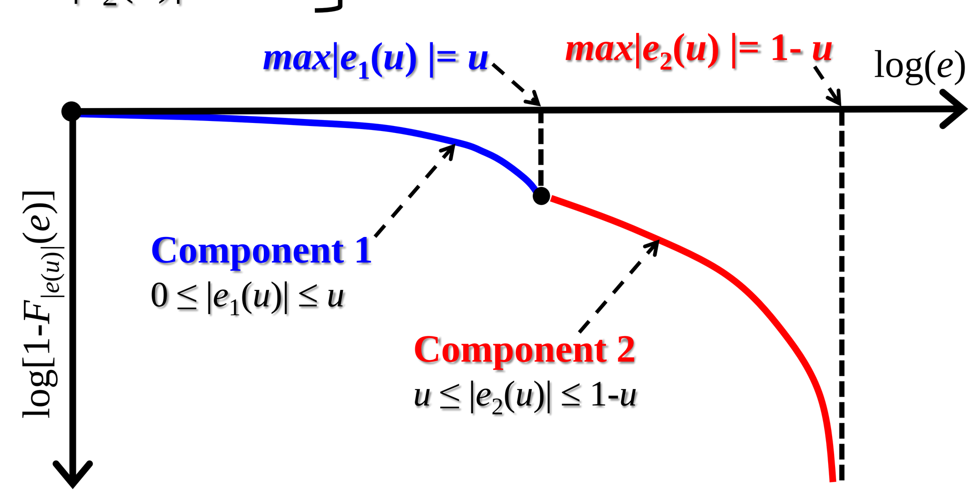

Figure 3: Schematic illustration of the components of the complementary cumulative distribution function (CCDF) of the standardized absolute error $|e(u)|$.

3. Statistical structure of hydraulic conductivity fields

Considering that hydraulic conductivity is **unknown** along the aquifer (see **Figure 1**), we use 4 stationary stochastic models to simulate scaling and non-scaling representations of hydraulic conductivity fields. Subsequently, based on the dimensional analysis presented in Section 2, we approximate the distribution of the standardized absolute error $|e(u)|$, using Monte Carlo simulations.

➢ ***Modeling hydraulic conductivities as a sequence of $r = L/l_c$ independent pulses of constant length l_c; see Figure 4.***

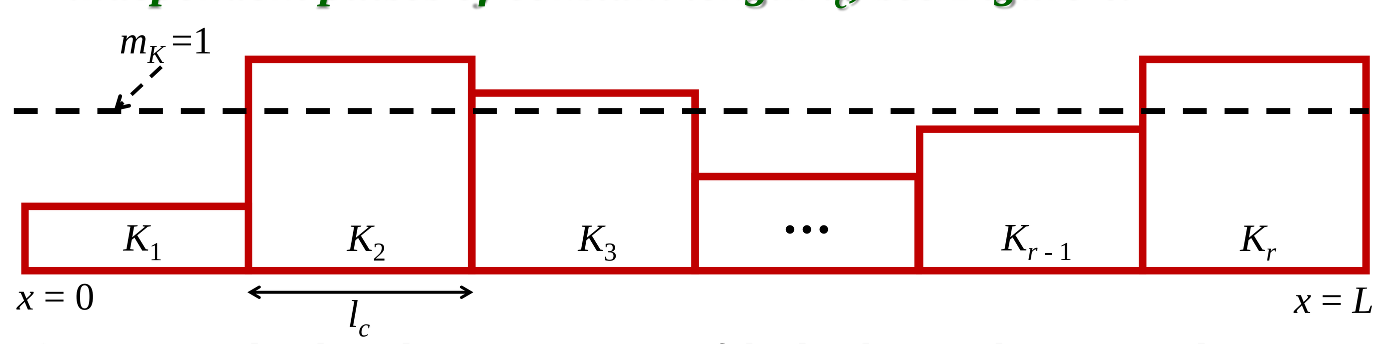

Figure 4: Pulse based representation of hydraulic conductivities along a 1D confined aquifer. Each pulse exhibits internal spatial variability.

○ ***Each stochastic model provides an alternative approximation to the dependence structure of the hydraulic conductivity field inside an independent pulse.***

A. Pulse based model

✓ Hydraulic conductivity is **spatially uniform** (i.e. constant) inside each independent pulse ➡ ***maximization of short-range dependencies***

- *K follows a mean-1 lognormal distribution with coefficient of variation CV_K (i.e. $K \sim LN(1, CV_K^2)$).*

B. Lognormal process with latent Markovian structure (LNM)

✓ Hydraulic conductivity is approximated by a mean-1 ***exponentiated discrete Markovian process***, with coefficient of variation CV_K. *The equivalent pulse length l_c is defined as the distance where the autocorrelation function equals 0.05 (i.e. 5%).*

↳ ***spatial dependencies also beyond the inter-borehole distance L***

C. Lognormal process with latent Fractional Gaussian Noise (FGN) statistical structure

✓ The hydraulic conductivity field is modeled as ***an exponentiated Fractional Gaussian Noise (FGN)***

➡ ***focus solely on long-range dependencies***

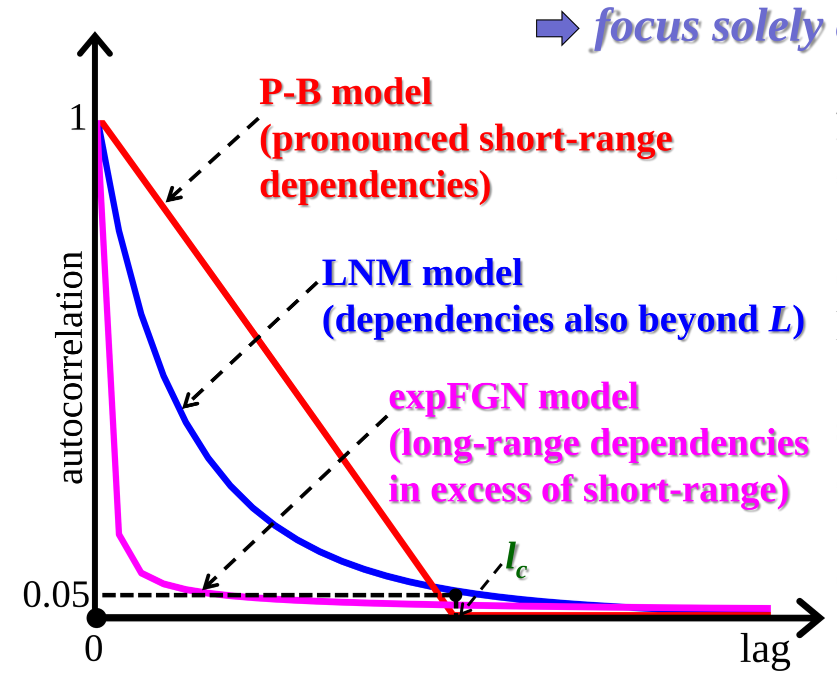

Figure 5: Schematic illustration of the theoretical autocorrelation functions of: (a) a pulse based (P-B) model (red curve), (b) a lognormal process with latent Markovian autocorrelation structure (LNM) (blue curve), and (c) an exponentiated Fractional Gaussian Noise (expFGN) process (pink curve).

D. Lognormal process with multifractal structure (MF)

Hydraulic conductivity is described by **the stationary stochastic self-similar process:**

$$K_l =_d A_s K_{sl}$$

where K_l is the spatially averaged hydraulic conductivity at scale $l < l_c$ (i.e. $s = l_c/l > 1$), $=_d$ denotes equality in all finite dimensional distributions, and A_s is a strictly positive ($A_s > 0$), log-infinitely divisible unit mean random variable.

Key parameters affecting the distribution of $|e(u)|$:

- ***u = x/L:*** standardized distance from the nearest measuring location (due to statistical symmetry; see Section 2.B and **Figure 2**)
- **Dependence ratio $r = L/l_c$:** a measure for the extent of apparent (i.e. observed) middle-scale heterogeneities in the aquifer. ⇒ **geologic maps**

4. Results – Conclusions

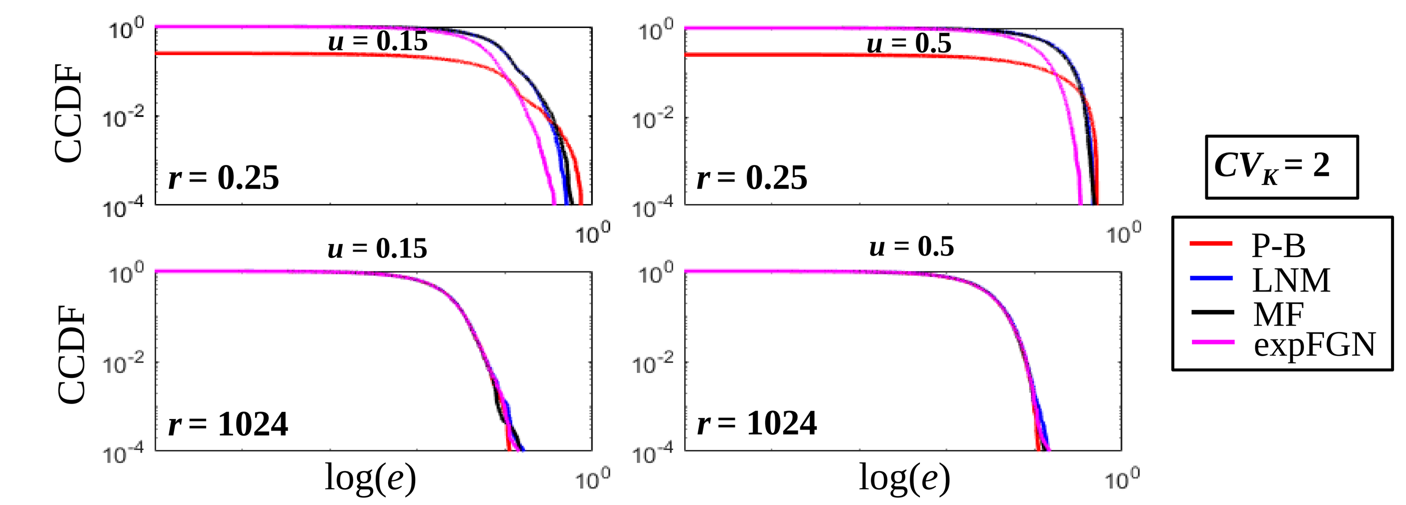

Figure 6: Distribution of the standardized absolute error $|e(u)|$, for CV_K = 2, dependence ratio $r = L/l_c$ = 0.25, and 1024, and standardized distances $u = x/L$ = 0.15, and 0.5. The corresponding curves have been obtained by ensemble averaging the results of 10000 Monte Carlo simulations, using models P-B (red lines), LNM (blue lines), MF (black lines) and expFGN (pink lines).

➢ For inter-borehole distances $L \le l_c$ (i.e. the characteristic linear scale of geologic formations in the study region), **the standardized absolute error $|e(u)|$ is underestimated by the pulse based (P-B) model (red curve)**.

➢ For $r = r_{max}$ **(complete independence)**, all four stationary stochastic models produce **same results for the distribution of $|e(u)|$**.

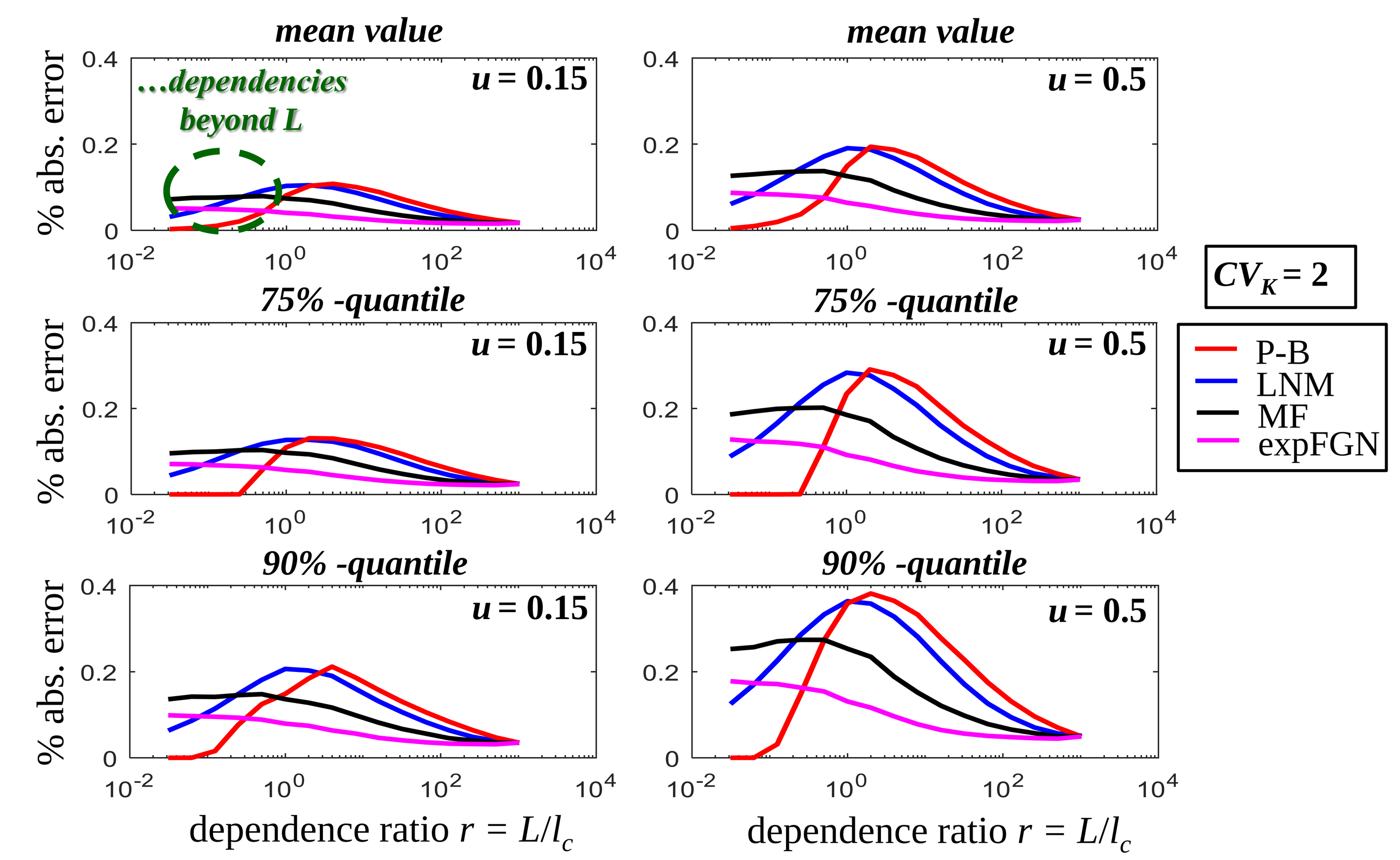

Figure 7: Mean value, 75%- and 90%- quantiles of the standardized absolute error $|e(u)|$ as a function of the dependence ratio $r = L/l_c$, for CV_K = 2, and standardized distances $u = x/L$ = 0.15, and 0.5. The corresponding curves have been obtained by ensemble averaging the results of 10000 Monte Carlo simulations, using models P-B (red lines), LNM (blue lines), MF (black lines) and expFGN (pink lines).

➢ For $L \le l_c$, dependencies at multiple scales dominate the distribution of $|e(u)|$. ➡ **LNM and MF models produce the maximum $|e(u)|$**

➢ For $L > l_c$, $|e(u)|$ decreases fast with increasing dependence ratio $r = L/L_c$. ➡ **At the limit as $r \to \infty$, the medium becomes statistically uniform**

Acknowledgements

This work is sponsored by the Onassis Foundation under the "Special Grant and Support Program for Scholars' Association Members".