Hydraulic head estimation at unobserved locations: Approximating the distribution of the absolute error based on geologic interpretations

Andreas Langousis, Vassilios Kaleris, Vagia Xevgeni and Foteini Magkou

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

1. Introduction

Assessing the availability of groundwater reserves at a regional level, requires accurate and robust hydraulic head estimation at multiple locations of an aquifer. To that extent, one needs groundwater observation networks that can provide sufficient information to estimate the hydraulic head at unobserved locations. Among other factors, the density of such networks is largely influenced by the spatial distribution of the hydraulic conductivity in the aquifer.

In this work, we study the distribution of the absolute error in hydraulic head estimation using: a) dimensional analysis, and b) two stationary stochastic models for simulation of hydraulic conductivity fields, with fundamentally different structures of spatial dependence: 1) a pulse based model with lognormal (LN) marginals, and 2) a discrete lognormal process with Markovian autocorrelation structure.

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.

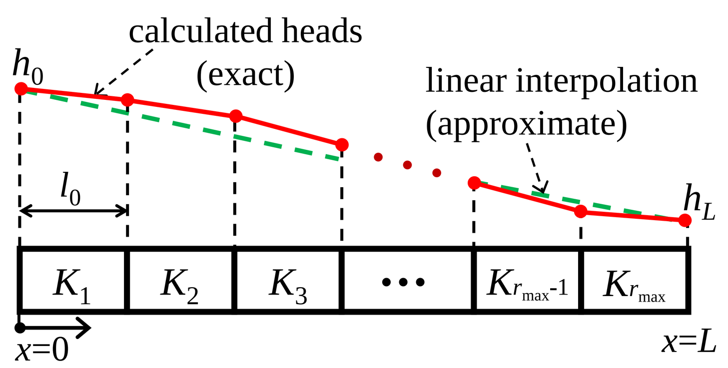

Figure 1: Schematic representation of the calculated (exact; red broken line) and linearly interpolated (green line) hydraulic heads in a 1D confined aquifer, formed by r_{max} successive hydraulic conductivity units of equal length.

In the case when the hydraulic conductivities K_i, $i = 1, 2, \ldots, r_{max}$ 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 $s = \mathrm{int}(x/l_0)$ 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:

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

For r_{max} = 4 and 8, **Figure 2** shows plots of the mean value $m_{|e(u)|}$ of the **standardized absolute error:**

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

as a function of $u = x/L$, assuming that k_i ($i = 1, 2, \ldots, r_{max}$) are independent realizations drawn from a lognormal (LN) distribution with unit mean value, and coefficient of variation CV_K = 0.1, 0.5, 1, and 2.

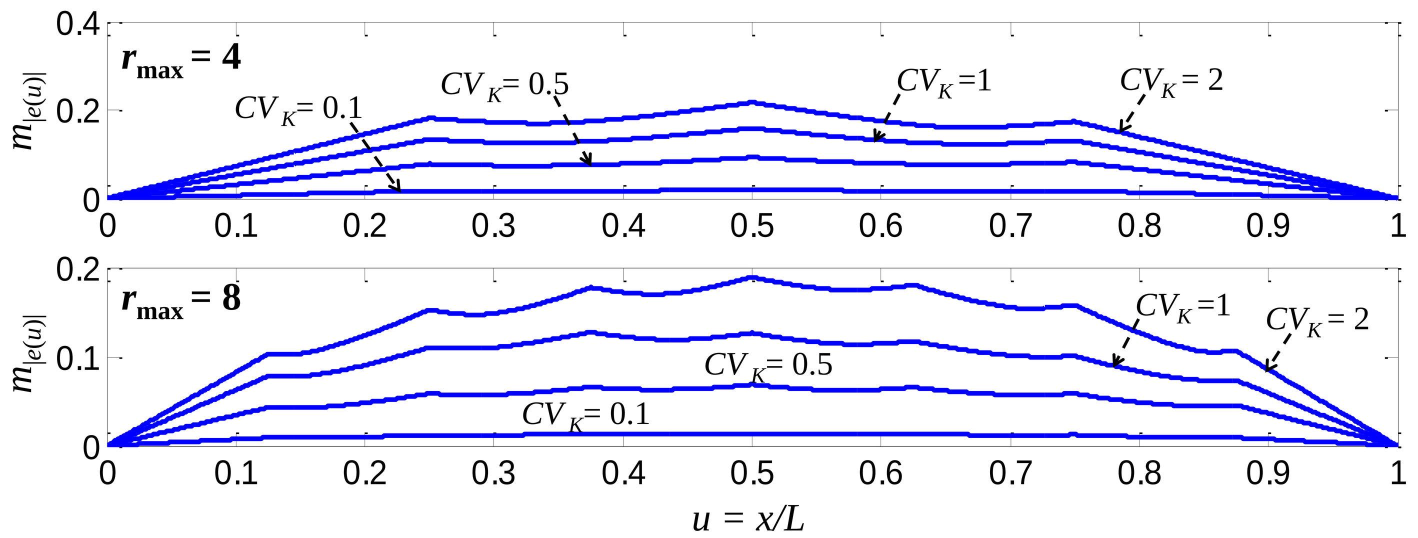

Figure 2: Mean value of the standardized absolute error $|e(u)|$, as a function of the standardized distance $u = x/L$, for different number of hydraulic conductivity units r_{max}= 4 and 8. The corresponding curves have been obtained by ensemble averaging the results of 1000 Monte Carlo simulations, assuming that k_i (i= 1, 2, …, r_{max}) are independent realizations of a lognormal (LN) random variable with unit mean value, and coefficient of variation CV_K = 0.1, 0.5, 1, and 2.

B. Theoretical attributes of the standardized absolute error distribution

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 3**):

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

- 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 4**):

Component 1: $0 \le |e_1(u)| \le u$
Component 2: $u < |e_2(u)| \le 1 - u$
for any $u = x/L \in [0, ½]$

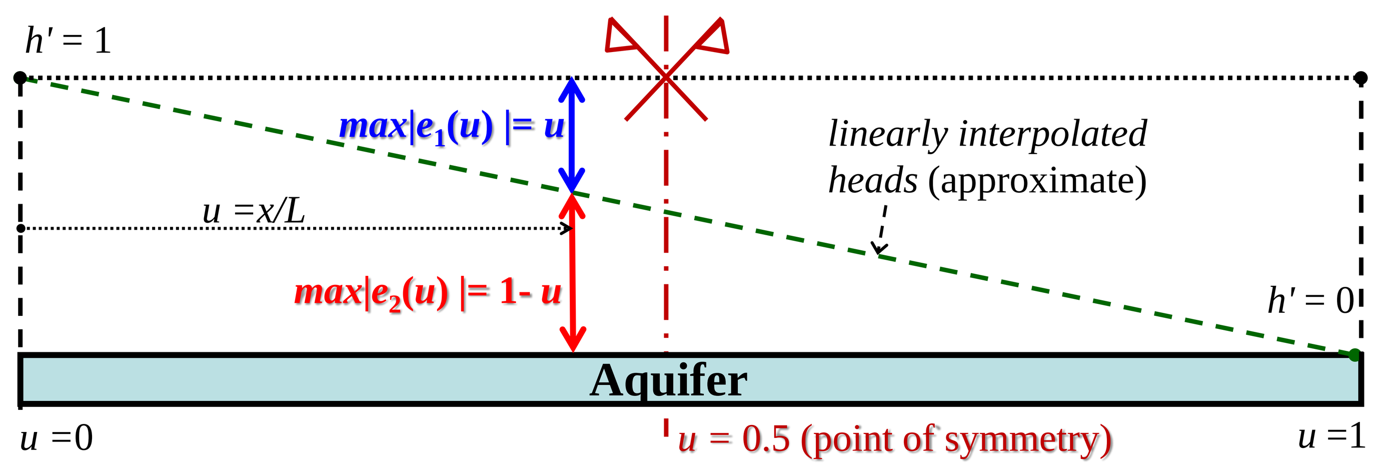

Figure 3: Schematic illustration of statistical symmetry in the study problem. $h'(u) = (h(u) - h_L)/(h_0 - h_L)$ is the standardized hydraulic head.

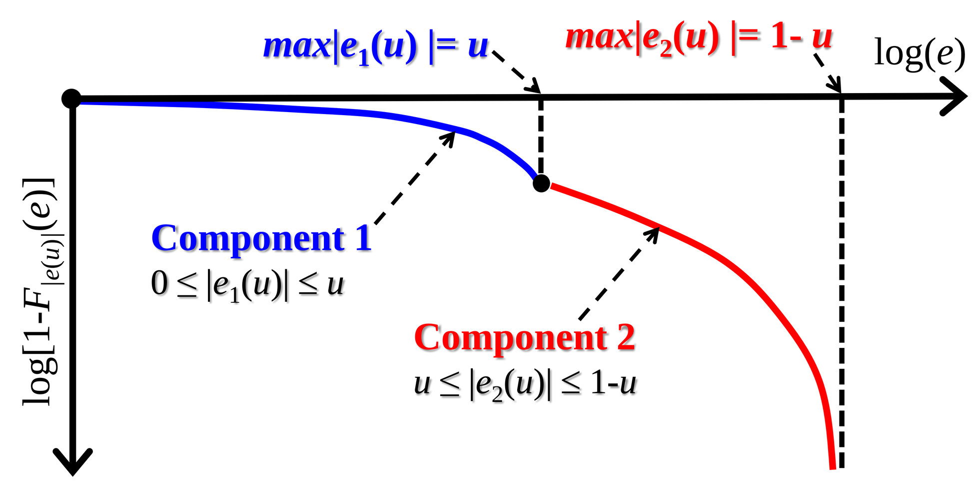

Figure 4: Schematic illustration of the components of the complementary cumulative distribution function (CCDF) of the standardized maximum error $|e(u)|$.

3. Stochastic modeling of hydraulic conductivities

Assuming **stationarity** of the hydraulic conductivity field at spatial scales much larger than the inter-borehole distance L, and based on the dimensional analysis presented in Section 2, we use two LN processes with different autocorrelation structures to approximate the distribution of the standardized absolute error $|e(u)|$, using Monte Carlo simulation.

A. Pulse based model ➡ *encompassing short-range dependencies*

In this representation, the geology of the aquifer between the two measuring locations is approximated by $r = L/l_c$ independent pulses of constant length l_c, with interfaces located randomly along the aquifer; see **Figure 5**. The magnitude k_i (i = 1, 2, …) of each pulse follows a mean-1 lognormal distribution (i.e. $K \sim LN(1, CV_K^2)$) with coefficient of variation CV_K.

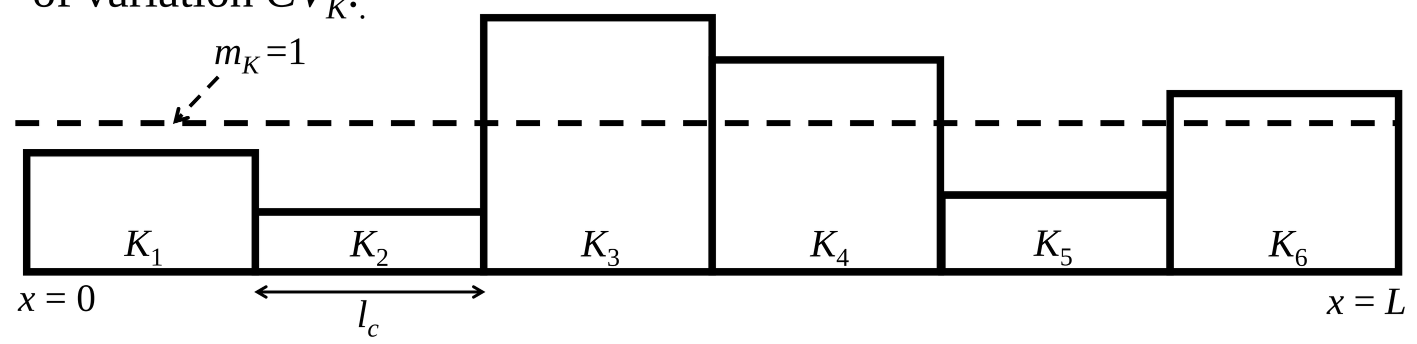

Figure 5: Pulse based representation of a 1D confined aquifer formed by r = 6 independent hydraulic conductivity pulses/units of constant length l_c.

✓ ***Due to complete dependence within different pulses, the effect of short-range correlations on $|e(u)|$ is maximized; see Figure 6.***

✗ ***Due to independence of different pulses, $|e(u)|$ is underestimated when $l_c \geq L$; i.e. for $r \leq 1$***

B. Lognormal process with Markovian structure (LNM)

The geology of the aquifer is approximated by *a discrete lognormal (LN) process with exponential autocorrelation function* (i.e. Markovian structure), and coefficient of variation CV_K. The equivalent pulse length l_c is defined as the distance where the autocorrelation function equals 0.1 (i.e. 10%).

✓ ***Accounting for long-range correlations, beyond the inter-borehole distance L; see Figure 6.***

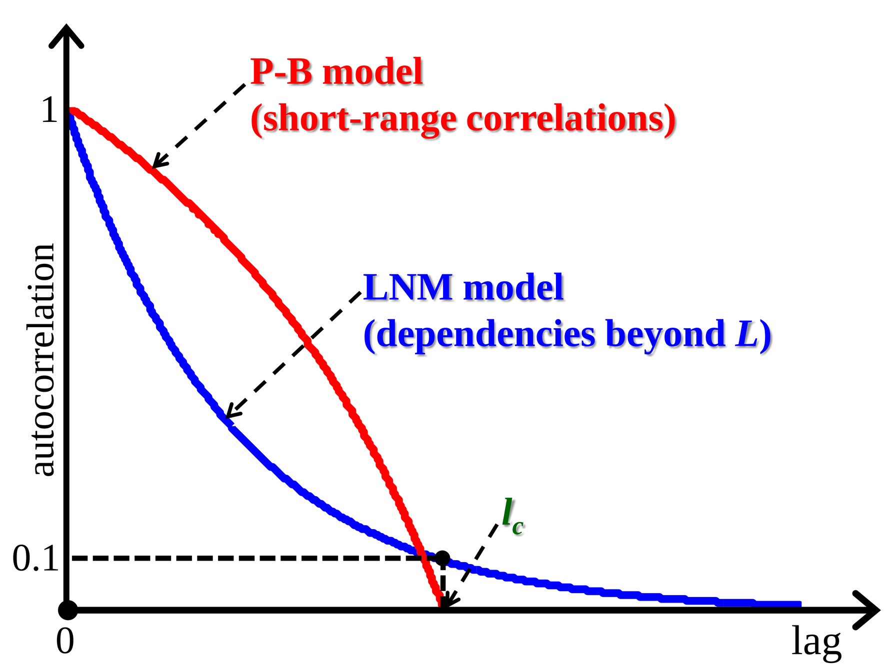

Figure 6: Schematic illustration of the theoretical autocorrelation functions of: (a) a pulse based (P-B) model with independent pulses of constant length l_c (red curve), and (b) a lognormal process with Markovian autocorrelation structure (LNM), and the same equivalent pulse length as that in (a) (blue curve).

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 3**)
- **Coefficient of variation CV_K:** a measure for the intensity of hydraulic conductivity fluctuations.
- **Dependence ratio $r = L/l_c$:** a measure for the extent of apparent (i.e. observed) middle-scale heterogeneities in the aquifer. ⇨ from **geologic maps** or preliminary in-situ investigations

4. Results

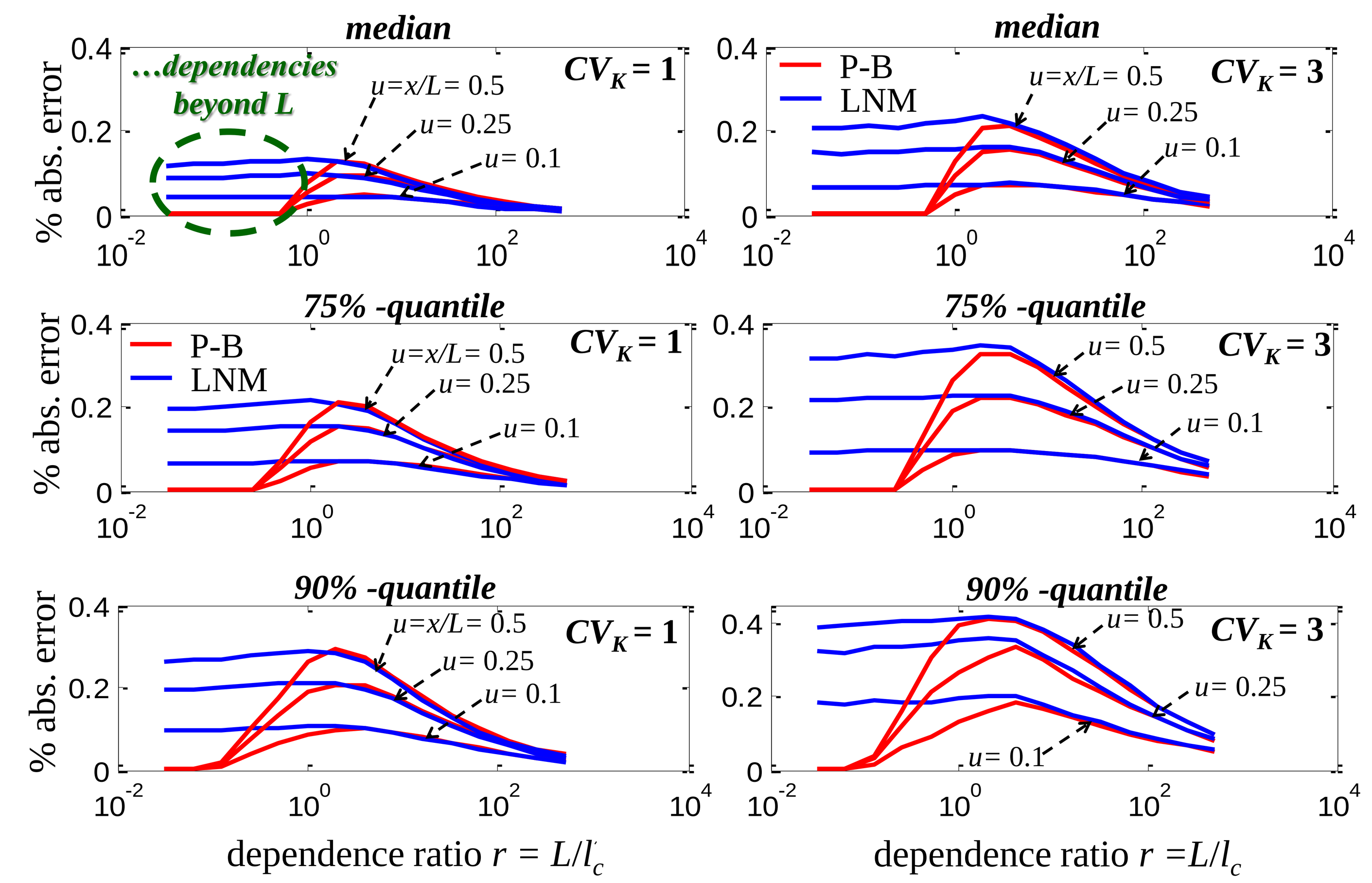

Figure 6: Median, 75%- and 90%- quantiles of the standardized absolute error $|e(u)|$ as a function of the dependence ratio $r = L/l_c$, for different values of CV_K = 1, 3, and standardized distances $u = x/L$ = 0.1, 0.25, and 0.5. The corresponding curves have been obtained by ensemble averaging the results of 10000 Monte Carlo simulations, using models P-B (red lines) and LNM (blue lines)

5. Conclusions

- The **standardized absolute error $|e(u)|$** in hydraulic head estimation **increases with the standardized distance u** from the nearest measuring location, and **maximizes at the middle** (i.e. u =0.5) **of the inter-borehole distance L**.
- **Mild dependence of the distribution of $|e(u)|$ on** the intensity of hydraulic conductivity fluctuations, as described by **CV_K**
  - ✓ ***Low requirements for detailed hydraulic conductivity information, based on laboratory samples***
- For inter-borehole distances $L \le l_c$ (i.e. the characteristic linear scale of geologic formations in the study region), **$|e(u)|$ becomes almost invariant to the dependence ratio $r = L/l_c$.**
  ➡ ***long-range correlations dominate***
- 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 uniform**

Acknowledgements

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