Department of Biostatistics and Department of Statistics
University of Washington

PHD APPLIED STATISTICS EXAM
September 1999

****Revised slightly for BIOST/STAT 578B, Spring 2001****

1 Background

Phreatic alluvial aquifers are widely used sources of water supply for a number of reasons. First, access to these resources is relatively easy, requiring only shallow pumping wells. Second, alluvial reservoirs are regularly renewed by rainfall infiltrations and local surface water. Third, from an economic point of view, they are generally located in the vicinity of the areas of heavy water consumption (urban, industrial and agricultural activities are frequently concentrated in alluvial valleys). Finally, the alluvial material itself, comprised mainly of sands and gravels, offers good hydrogeological characteristics.

Unfortunately, some of these advantages are also balanced by vulnerability factors. First, the proximity of human activities is also the source of various risks. Pumping areas, formerly located in natural environments are now situated in the closest neighborhoods, and sometimes the very heart of cities as a result of growing urbanization. Second, the degradation of the global quality of surface waters, which often represent an important component of the balance of shallow groundwater resources (aquifer replenishment) has now become an actual threat to the quality of the latter. Third, the lack, in most cases, of significantly thick and wide upper geological layers of low permeability (natural barriers), increases the potential impact of pollutants that may accidentally infiltrate.

The problem of predicting impacts, quantifying risks and estimating the degree of vulnerability of such a pumping field is particularly crucial when hydrogeologists rely, for expertise and eventual recommendations, on mathematical conceptual models. The unavoidable simplifications in the geological features of the porous material, as a result of both the limitations of the computer models and the lack of field data, are particularly critical in the case of alluvial aquifers, for the reasons noted above. Fortunately for practical studies, the discrepancies between model calculations and field observations, at least regarding transfer times and concentration levels, tend to decrease in inverse proportion to the distance between the pollution source-term and the impact point. This fact is a consequence of the averaging and smoothing of small-scale heterogeneities in the continuum representation of the porous material. Nevertheless, these deviations may become important in cases when one has to predict the impact of nearby pollutions or accidents, e.g. for vulnerability assessment. In that case, a precise description of the geology is needed, with increased costs of investigation.

The general objective of this project is to show how data reflecting the response of a well to a nearby pollution source can be effectively studied and modeled using mainly exploratory statistical tools. This inverse approach may lead to a more reliable undertanding of (i) the local characteristics of the geological terrain on transfer times and (ii) the factors that may influence the subsequent pollution of the pumped water (pumping rate, hydrology, etc.).

2 The Data Site

The case study on which this analysis is to be based is the main pumping area of a large urban community. This water supply facility, which accounts for 60% of the local water needs, consists of eleven wells distributed over 156 acres of wooded land in the southern suburbs of the city and close to a major river. The relative locations of the wells in this pumping area and their relationship to the river are depicted in Figure 1.

Well mechanics show that the contribution of a given well to the global output of the combined network of wells (q) remains constant, regardless of the pumping rate. Calculated contributions of each of the 11 wells are presented in Table 1. These calculations show that four major wells, W6, W1, W4, and W18, provide almost half of the total output. The impact of a contamination affecting one or several of these wells is of particular importance.

The neighboring river has both a quantitative and a qualitative impact on the well production of this site. First, it represents a major term in the local water balance of the aquifer, as the latter is directly replenished by the stream water, and second, as a consequence, the poorer the quality of the river, the greater the vulnerability of the pumping field.

The river is not the only factor of risk for this area. Many other terms deserve special attention, including: inland industrial activities, minor streams, sewers, and former gravel quarries now filled with water. But the vicinity of the river, its contribution to the water balance (given in Table 1), and the frequency of pollution events makes it a priority to assess its relationships with the wells.

One important parameter is the volumetric contribution of the river water to the individual balance of each well, which leads to a complementary classification of the wells, in terms of the potential impact of a pollution event in the river. By river water, we mean the water infiltrated into the aquifer *at the regional scale* and reaching the pumping well.

Theoretical model calculations of the local water balance show that the river contributes more than one third of the global production of this field of 11 wells. However, this average proportion varies notably from one well to another, due to the relative position of the wells to the river and other terms of the balance. The wells closest to the river are expected to depend most on the river water. Table 1 reports the detailed results from the steady state modeling of the regional water balance. One can see that the estimated proportion of river water ranges from 8.5% (in well W1) to 76.1% (in well W6). Four wells exhibit contributions above 50% : W6, W5, W9 and W8. As expected, these wells are the closest to the river. One important observation is that well W6 appears at the top of both classifications : productivity and potential impact of a pollution in the river. The study of this specific well is thus of particular importance.

Table 1		
	q	r
W1	12.0	8.5
W2	6.5	17.2
W3	8.8	16.9
W4	12.0	14.2
W5	8.2	70.1
W6	14.1	76.1
W8	7.3	51.9
W9	6.8	55.2
W16	5.8	15.9
W17	7.9	30.1
W18	10.3	46.8

q: Individual contributions of wells to the global output.
r: Theoretically calculated percentage of river water in the output of the individual wells.

3 Problem Summary

Data are available to study possible relationships between chloride concentrations in 7 (of the 11) wells and chloride concentrations in the river. It is expected that there might be some delay between the time that changes in concentrations are noted in the river and the time that any effect of those changes might be noted in the well samples. This delay or time lag is essentially the time for transport of water from the river through the aquifer to the well. Chloride concentrations in the river may vary slowly over time within a year or from year-to-year; these are not the variations of interest from the perspective of the well concentrations over the 2-year time period for which data are provided (see below). Mathematical models for the transfer of a variable input source-term into a quasi-stationary flow velocity field suggest that transfer times are on the order of 100's of days, depending on the particular point along the waterway for the calculation, and varying greatly from one well to another.

Other factors to be considered in modeling the well chloride concentrations are the river flow as given by daily river height measurements, the rate at which water is being pumped from the system of wells, and rainfall infiltration as provided by a calculation derived from daily precipitation data. Higher water levels might be expected to induce higher river infiltration rates, and hence, higher well chloride concentrations, but there is little theory to suggest what time lags might be relevant for the possible effects of variations in river height. Some mathematical arguments have been proposed for how a transient well chloride signal follows an increase in pumping rate, but these are not well-determined.

4 Database Definition and Access

(a) **WELLCONC7**: 96 nearly weekly measurements of chloride concentrations (mg/l) in 7 of the 11 wells pictured in Figure 1: W4, W5, W6, W8, W9, W17, W18. Measurements were typically (but not always) made on Mondays, with the exception of the Christmas-New Year holiday periods. The dates of the well measurements, from 89/01/09 to 90/12/27, are provided as an Splus "dates" object, **WELLCONC.date**.

(b) **RIVCONC**: daily measurements of river chloride concentrations (mg/l) spanning 12 years, 79/01/01 to 90/12/31 (**RIVCONC.date**). (Note that the full length of this series of measurements is not relevant to the modeling of two years of well chloride concentrations, but it may be useful for understanding or characterizing variation in river chloride concentrations.)

(c) **RIVHGT**: daily measurements of river heights (meters) spanning 5 years, 86/01/01 to 90/12/31 (**RIVHGT.date**).

(d) **PRATE**: daily measurements of the output of the pumping station for this well field (cubic meters/day/1000) for six years, 85/01/01 to 90/12/31 (**PRATE.date**).

(e) **INFIL7**: mathematical model-based calculations of rainfall infiltration for the two-year period of well chloride measurements. These have been reduced to 7-day cumulative infiltration figures for the 7 days preceding each well measurement.

These data will be available to you as Splus objects saved in an ascii (text) file created using the Splus "data.dump" command. The data are not now available, but after class next week (April 5), you can copy this file ("wellconc7.dump") from the class web site and then use the Splus "data.restore" command to restore these Splus objects into an Splus .Data directory.

In Splus type:

```
data.restore(``wellconc7.dump'')
```

.

5 Task

You are asked to determine if there is a statistically identifiable response in (any of) the wells to fluctuations in chloride concentrations in the river, taking account of the influence of the four principal covariate series, (b)-(e), above. Time delays for the possible response of wells to the river must also be considered. However, although these delays depend on "distance" to the river, both the path of the river around the pumping field and the complexity of the aquifer hydrology complicate this simple notion, so no measures of this distance are provided. The general layout given in Figure 1 and the information provided in Table 1 may be considered in judging or interpreting your empirical models.

Summarize your findings in a report suitable for an applied scientist. Comment on whether and/or how the model(s) you have found might be used for predictive purposes by water managers in the case of a pollution event in the river. Comment on any limitations that come to mind regarding the use and interpretation of the model(s).

Figure 1.

The relative positions of the 11 wells, the river, and connected waterways. The river runs essentially north-south with "C" marking the point on the banks of the river closest to the area of the wells. Points "A" and "D" are located on waterways (basins) connected to the river.