DYNAMIC NETWORK DESIGN FOR ESTIMATION OF GROUNDWATER FLOW AND MASS TRANSPORT IN A ONE-DIMENSIONAL AQUIFER USING THE KALMAN FILTER

A THESIS
SUBMITTED TO THE DEPARTMENT OF APPLIED EARTH SCIENCES AND THE
COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

By
José Jaime Gómez Hernández

March 1988

Approved for the Department:

Adviser

Approved for the University Committee on Graduate Studies:

Dean of Graduate Studies

ACKNOWLEDGMENTS

I wish to express my gratitude to several people who have helped me to write this thesis and to survive its writing. Thanks to Professor Irwin Remson for his full dedication to his students, and in particular to me. He opened the doors for me to get into Stanford University and shortly afterwards he opened the doors of his office, his house; he offered unconditional support and help and was the person always ready to help if any problem arose. Thanks to Professor David Freyberg for his comments and suggestions throughout this study, and for the ones to come in our on- going research. Thanks to all my friends who encouraged me to keep moving in the moments of depression, and in particular thanks to my wife, Inma, who had to support my bad temper when the computer was down and my typing late at night while she tried to fall asleep, and without whose support and care I would have not survived for two years in a strange land far from our beloved Spain. Thanks also to my parents, for whom, I know, these have been two very long years, and to the rest of my family who always cared so much and backed me up in all my goals.

I would also like to gratefully acknowledge the Generalitat Valenciana of Valencia (Spain) which funded my stay at Stanford through its Becas de Estancia en el Extranjero para Investigadores de la Comunidad Valenciana and to the Joint Committee for Scientific and Technological Cooperation between the USA and Spain which partially funded this project under grant No CCA830917.

Finally, I cannot end without acknowledging my Macintosh com¬puter which did all the routine work and which, unfortunately, took too many hours from my time, at the expense of spending it with my family and friends.

ABSTRACT

Network design for estimation of groundwater flow and mass transport in aquifers must include the time component so that not only where but when the sampling should be taken is determined. For this purpose, standard linear estimation techniques for static problems are coupled with variance propagation in time through the use of the Kalman filter to devise a methodology for dynamic network design.

The criterion used to discriminate between possible sampling schemes is the mean estimation error variance over time and space for the entire simulation period and the entire aquifer. Shorter spans of time and/or parts of the aquifer could have been used instead. The optimization is constrained by the location of the possible sampling points (in both time and space) and by the total number of samples to be taken. For the results presented here the optimization is carried out by exhaustive search, which results in a very expensive method.

The solution to the network design problem for estimation of groundwater flow and mass transport in a one-dimensional aquifer with uniform parameters permits one to draw some conclusions that may ease the problem for more realistic higher-dimensional problems. In groundwater flow the optimal location is mainly related to the geometry of the system. However, no relationship between the optimal sampling time and the system parameters is found. In mass transport the timing is determined by the travel time, and the location by the geometry and the boundary conditions. The Peclet number seems to influence only the level of uncertainty, not the design.

TABLE OF CONTENTS PAGE

LIST OF FIGURES PAGE

1. INTRODUCTION

The problem of network design has been recognized since the very moment that somebody had to decide where to take a number of samples to establish the state of a system. Initially, the good judgment of an expert was used to decide where and when to take the samples, but the cost of sampling and the need to optimize the information provided by those samples lead to the problem studied here: to find the sampling location and times such that those samples are optimal under a set of criteria.

Network design for phenomena that do not vary in time or vary very slowly has been developed in a number of fields. But I would like to include time in the network design problem so that phenomena that clearly vary in time, as well as in space, can be addressed. Some attempts have been made in this respect but they reduce the problem to a static one by separating space and time. In groundwater hydrology, a feasible methodology for dynamic network design has not been found to apply to practical problems. The dimensionality of the problem is one of the major drawbacks. Notice that a small three-dimensional aquifer with 10∞10∞10 cells requires a 1000∞1000 dimensional matrix to represent all the two-point covariances that theoretically must be propagated in time and heavily manipulated (see Chapter 4) in order to include time in the network design problem.

The methodology presented here is applied to a uniform one-dimensional aquifer in order to try to get some understanding of the network design problem, with the goal of easing the solution of more complex systems. It is based on the following hypotheses: i) the aquifer parameters are exactly known; ii) the external stresses, if any, are also perfectly known; iii) the initial state of the system is either completely unknown or known with a given estimation error covariance; and iv) measurement errors are allowed.

Since the system is known with imprecision at the beginning of the simulation period, this imprecision will propagate in time. The objective of the network design problem for this study will be to find the sampling scheme that most reduces the mean estimation error variance

over time and space. Many other objectives can be chosen, as discussed elsewhere in this report.

The procedure consists of the following steps: i) define a set of possible sampling locations and times; ii) set the total number of samples to be taken, iii) for each of the possible ways to take as many samples as stated in ii) among the locations set in i) propagate the error covariance from time zero until a measurement is made, propagate the new filtered error variance until another measurement is made, and so on until the end of the simulation period, iv) calculate the mean estimation error variance over space and time; v) the sampling scheme chosen will be the one with the smallest mean estimation error variance.

The work is organized into five chapters. Chapter 2 reviews the work done so far in the field of network design. Chapter 3 derives the transfer matrix of the state-space models used in groundwater flow and mass transport. Chapter 4 presents the formulation of the linear Kalman filter. And Chapter 5 presents the results for a one-dimensional aquifer for both the flow and mass transport problems.

2. LITERATURE REVIEW

There are several ways to classify the studies that have been done in network design. Maybe the two most logical approaches are to classify them either according to the objective of the network or according to the procedure used to pursue that objective. In the following paragraphs a review of the most relevant work in network design will be presented using the latter approach.

Since the late sixties considerable research has focused on the issue of network design. Optimal allocation of sampling points in both time and space has been the primary goal. The first major problem arises in the definition of optimal. Who defines it? What are the criteria used? In many instances a decision maker is needed to provide a subjective evaluation of such factors as the cost of a unit of estimation error variance, or the amount of risk that users can afford, or to set priorities in the case of multicriterion objectives, or to state social and economic

constraints on the optimization problem. Therefore, the network design problem can require inputs from a variety of fields, in many cases from outside the hydrologic setting. This study does not take into account such a decision maker and optimal is defined in terms of the mean estimation error variance. However, there would not be any problem in extending the definition of optimal to include, for instance, the cost of sampling. Several definitions of optimal and the need for a decision maker's subjective evaluation will be seen in some of the scenarios presented hereafter.

The different studies that will be reviewed can be divided into two groups depending on whether or not the dynamics of the phenomenon under study are included in the design of the network. When the phenomenon is constant in time or varies so slowly compared to the sampling frequency that it can be considered as time-invariant, the network design will be said to be static. When, on the other hand, the phenomenon varies in time at a rate similar to the expected sampling frequency, the dynamics of the system must be included in the design problem and the network design will be said to be dynamic.

Many are the fields of application and many are the techniques that have been used to solve the design problem. Likewise, the objectives are countless, but in essence a network is designed to gather data at a number of points with the objective of obtaining local or global estimates of a given variable within a given region. Whether these estimates will be used to enforce a policy or to calibrate a numerical model will determine the final goal of the network. My major emphasis, in the review that follows, will be focused on applications that have been carried out in hydrology, spanning from surface hydrology to groundwater hydrology, for both static and dynamic network design. Some of the techniques used are geostatistics, sampling theory, linear systems, state estimation theory (mainly the Kalman filter), information theory and others. The presentation will be divided into two parts, one dealing with static network design and the other with dynamic network design.

2.1. Static network design

All the work presented under this heading is characterized by regarding the phenomenon under study as time-invariant or slowly time—variant so that no temporal evolution is taken into account.

Geostatistics has been the most used tool in static network design. The theory of regionalized variables and the method of kriging developed by Matheron [1971] have been successfully applied in hydrology and, in particular, in network design. Given a region, and a number of points where the values of the variable under study are known, kriging [Journel and Huijbregts, 1975] provides an estimate of the variable at unknown points as a linear combination of the known values. Spatial correlation is included through the variogram, which is a measure of the expected variability between points. The coefficients of the linear combination are computed such that the estimate is unbiased and minimizes the variance of the estimation. For this reason, kriging is called a best linear unbiased estimator, or BLUE. The fact that the kriging estimation variance as well as the coefficients of the linear combination are independent of the actual values at the sampling locations makes kriging particularly powerful in network design since no measurement must be done in order to know how the estimation variance will change by adding a new sampling location.

In the geostatistical approach, the process is assumed to be second-order weakly stationary. This means that the expected value must be constant over all the domain (under certain circumstances this assumption can be relaxed [Delfiner, 1975]) and that the expected value of the square of the differences between two points must be dependent only on their relative position.

Delhomme [1978] showed the potential of geostatistics to be used in network design for both global and local estimation. In the ouadi Kadjemeur basin (Chad), global estimation of mean areal rainfall was done from a set of 33 rain gages. The fictitious point method was used to determine where the 34^{th} gage should be located. The fictitious method consists in the selection of that point that provides the maximum reduction on global estimation variance. The

name of fictitious comes from the fact that this reduction can be computed without actually installing the gage. This is possible because the kriging variance is independent of the data values themselves. In general a map of isoreduction of variance curves accompanies this kind of studies in order to detect not only the maximum reduction but also the zones that need to be reinforced.

For local estimation, Delhomme presented the case of estimation of hydraulic heads in the Huveaunne aquifer in France. In this case there are as many kriging variances as points in which the hydraulic head is to be estimated (in the global case only one estimate of the mean areal value was wanted, therefore there was only a single estimation variance). Therefore, the inclusion of an additional gauge will influence the estimation variances at every other point. One may choose between selecting that point that provides the maximum reduction in the areal integral of estimation variances, or simply locate the new stations at the local maxima of the initial estimation variance map.

Jones et al. [1979] used ordinary kriging for network design for estimation of rainfall totals during a particular time interval. A very sophisticated method was proposed to estimate the parameters of the covariance function from a number of data values. To obtain the kriging estimation variance (or mean square error as it is called in the study) the zone is divided into a square grid and a different covariance function is computed for each square using a surrounding square of larger dimensions. With the computed covariance and the gages within the outer square, the interpolation procedure is carried out for the inner square and the estimation variance is computed in the inner square. Repeating the process for each square, a map of estimation variances is obtained. The procedure to reinforce the existing network is simply to add gages where the estimation variance is below the accuracy required by the users.

Villeneuve et al. [1979] used ordinary kriging to reinforce the Quebec streamflow recording network in Canada. After a first screening of the existing network in order to eliminate those clearly superfluous gages, the problem of network reinforcement was addressed for local and global estimation of two variables: maximal 1-day flow values with a

return period of 100 years and minimal 7-day flow values with a return period of 100 years. Although the mean areal values of these variables do not have much meaning, the fictitious point method was used to locate the gage that produces the maximum decrease in the estimation variance of these mean areal values. The local estimation problem was more interesting. With the objective of reducing the kriging estimation variance as much as possible in as many points as possible, a number of fictitious stations were included at once and the new map of kriging estimation variances was compared with the initial one. This process was repeated until those stations (among a set of 10) that produced the most significant variance reduction over all the area were found. Different sets of gages proved to be the best choice for the two variables studied (maximal flow and minimal flow), raising the problem of multicriterion network design. In order to finally decide which is the best set that simultaneously takes into account the two estimation problems, some sort of weighting of the estimation variances must be introduced by the decision maker.

The initial work by Matheron [1971] was extended by Delfiner [1975] in order to deal with processes that are stationary only after an incrementing transformation. Such a process is called an intrinsic random function (IRF) and the correlation structure is represented by a generalized covariance that coincides with the process covariance if the process is stationary, but that can be interpreted only in connection with the incrementing process. This theory is used by Hughes and Lettenmaier [1981] for linear estimation and network design. The study was carried out in the Spokane Valley aquifer in Washington State and the variable of interest was nitrate concentration. Sixty-two wells that constituted the existing network were used to identify the underlying IRF. The criterion used to reinforce the network was to minimize the variance of an areal estimate of the nitrate concentration. The method proposed goes beyond the fictitious point method in that it tries to allocate a number of stations at once instead of only one at a time. For this purpose, the property that the kriging estimation variance depends only on the generalized covariance and on the relative geometry of the samples but not on the observations is used to write the kriging estimation variance as a function only of the sampling locations. A first-order Taylor's expansion of the kriging estimation variance as a function of

the coordinates of the stations to be added was used as the objective function. The minimization was carried out using a gradient method and the coordinates that allocate the stations with minimum kriging variance were obtained.

McBratney et al. [1981a,b] used ordinary kriging to design a network for sampling of soil pH and soil thickness. Assuming that a semivariogram is known from a previous survey or from any other source, the authors devised a method to obtain the optimal sampling using a regular sampling pattern that minimizes the maximum point or block estimation variance. Since no preexisting sampling data were considered, it is straightforward to conclude that, for isotropic variation, the best sampling pattern corresponds to an equilateral triangular grid and the maximum point estimation variance happens at the center of the triangle. In the case of block kriging the maximum estimation variance can occur either when blocks are centered at the center of grid cells or when they are centered on observation points, so that both cases should be solved to identify the best sampling procedure. The kriging variance can then be computed and plotted for different distances between sampling points. From the resulting graph of network density versus precision, the density may be read if the precision is the decision variable, or viceversa. A transformation of coordinates is suggested to deal with anisotropy. Two examples are presented, one to design a network for measurement of the pH of topsoil at Suffolk, England, and the other to design a network for measurement of soil thickness at Lephinmore, Scotland.

In the last work of McBratney et al. [1981a,b], as in the work cited in the following paragraphs, the decision variables are not sample locations but just one or two variables characterizing a regular network. Besides the fact that the proposed methodologies do not take into account any preexisting stations, they cannot be applied to domains with irregular boundaries or sizes of the order of magnitude of the sampling interdistance since the expression for the estimation variance is computed under the assumption of an infinite domain.

A very similar method to that of McBratney et al. [1981a,b] was used by Sophocleus et al. [1982] and Sophocleus [1983]. In this case, the authors used universal kriging to solve a

network design problem in the Ogallala formation in northwestern Kansas for measurement of groundwater levels. The issue is not to reinforce a network but to find which uniform pattern and with which density the sampling should be done to reduce the kriging estimation variance at a reasonable cost. The only decision variables are the sampling pattern (i.e. square pattern, triangular pattern, etc.), the density (generally measured as the interdistance between adjacent samples) and the orientation. For the case study contemplated by Sophocleus, the square pattern was chosen from political considerations. Groundwater head variation, after a linear drift was removed, appeared isotropic, so the orientation did not have any influence on the final solution. Thus, the only decision variable was the network density. To determine the optimal network density, the kriging estimation variance was evaluated at the center of the square for different density values. The trade-off between the reduction of kriging estimation variance and the cost of the stations needed for that purpose was left to be set by the decision maker. The results showed that a network consisting of a well every four miles yields regional groundwater maps with the same uncertainty as the current network. The results also showed that to reduce that error by $^1/_2$ the sampling effort would have to be increased by 16 times.

Carrera et al. [1984], also using a geostatistical approach, employed a branch method to obtain the optimal location of a number of measurement points from a discrete set of possible sampling points. Only Hughes and Lettenmaier [1981] have tried to find the optimal location of a number of sampling points at once, but since they needed to linearize the expression of the kriging estimation variance to apply their optimization algorithm, their solution was sub-optimal. Thus, the work by Carrera et al. is the first one in which kriging is used to find the optimal location of a number of stations. It is applied to the optimal selection of measurement points in the San Pedro River basin in Arizona to estimate fluoride concentration of the groundwater. The optimization problem consists in selecting wells among m possible candidates such that the estimation variance of the mean value of the fluoride concentration over the entire basin is minimum. A branch method is used. The algorithm starts computing the estimation variance as if a measurement were to be done at each and every possible point and proceeds discarding measurement points and recomputing the estimation variance until the

optimal solution is found. Some properties of the estimation variance make it unnecessary to perform an exhaustive search through all the branches and a procedure is given for an efficient update of the estimation variance when a point is dropped. A sensitivity analysis on the variogram parameters (range and nugget effect) was carried out showing that the results were very robust. The same optimal configuration was found in almost all cases and when it was not, the new configuration only differed in one station.

Bogárdi et al. [1985] combined multicriterion decision-making and geostatistics to design a regular observation network for estimation of several parameters. The network pattern was rectangular, not necessarily square, thus accounting for possible anisotropy. An expression for the average point estimation variance over one of the rectangles, as a function of network density, orientation, average distance and variogram was obtained using the geostatistical formalism. One such expression was obtained for each of the parameters to be estimated, the only difference between expressions being the particular variograms of each parameter. Although the average point estimation variance over a rectangle was used, the authors acknowledge that average point (or block) estimation variance over the whole area or maximum point estimation variance, could have been used. Afterwards, composite programming was used to minimize an objective function that included statistical criteria, observation effort criteria and their corresponding trade-offs. The statistical criteria are related to the estimation variance. The observation criteria are reduced to one criterion for which the network density acts as a surrogate. Trade-off coefficients are needed, both among statistical criteria (without directly measurable economic consequences) and between statistical criteria and the observation effort criterion. The objective function is finally minimized using a gradient optimization. The two major drawbacks of this method are: i) it assumes that the variograms for all the parameters to be estimated can be inferred from prior regional information or data from part of the area and ii) it is assumed that there are no operating stations already in the area. In addition, the subjective problem of assigning values to the trade-off coefficients appears here as it appears in any multicriterion decision problem. For the case in which stations are already operating in the area under study, a procedure to adjust the best network previously

obtained for the existing stations is suggested. In this procedure the network obtained by the unconstrained method (no preexisting stations) is superimposed to the existing network. This determines the number of stations that should be added and gives an initial location that will be used in a gradient optimization method with the coordinates of the new stations as decision variables. The method was applied to design a network to identify four aquifer parameters: thickness and porosity in a two-layer aquifer in western Hungary. The variograms were estimated from 37 samples taken in a small subarea of the aquifer (500 x 250 m) and an unconstrained network search was carried out.

The last work approaching static network design from the geostatistical standpoint that will be cited here is the one by Chou [1986]. In this work non-linear programming and geostatistics were used to select the optimal location of a number of stations at once. The decision variables are the sampling point coordinates. The statistical criterion used to compare different configurations is the variance of the mean estimate of one parameter over the whole domain. Therefore the network is designed for global estimation. An expression of the estimation variance is obtained in terms of the coordinates of the existing stations and in terms of the unknown coordinates of the additional stations, then an algorithm that uses a simplified version of the self-scaling gradient method, the steepest descent method and the PARTAN ridge following technique is used to obtain the optimal location of the additional stations. A synthetic case is used to study the feasibility of the method. The results for this example are compared to the results that would have been obtained by sequential application of the fictitious point method. It is my judgment that neither the changes in sample locations nor the reduction in variance was large enough to justify the additional effort required by the proposed method.

At this point, after the presentation of the most relevant papers in spatial design of networks for sampling time-invariant phenomena, it is worthwhile to point out one remark made by Switzer [1979] about optimal design of spatial networks. The minimization of the mean square error or kriging estimation is a very difficult exercise from the mathematical point of view. Optimization becomes wasteful since good judgment on the part of the designer will

produce a suboptimal configuration as good, for practical purposes, as the optimal solution. Good judgment means avoiding clusters of stations, taking into account anisotropy, trying to spread stations at a distance about the correlation length of the process, etc. By doing so and applying simple weighting schemes, the network can be used to obtain nearly optimal estimates.

Geostatistics has been shown to be a very suitable tool for network design, especially for estimation of time-invariant processes. However, there have been authors that have used other techniques for the same purposes. Among these techniques, that have not had a broad application, I will review three works. One work based on linear systems theory, another work using a formalism essentially equal to the geostatistics approach, and finally a different method using information transmission theory are discussed below.

Eagleson [1967] used deterministic linear systems theory in the frequency domain [Bracewell, 1978] to determine the optimal density of rain gages required to accurately forecast discharge . The catchment was modeled as an infinite array of time—invariant independent linear subsystems. The peak discharge at the outlet was the variable to be estimated. A linear system related precipitation and discharge through an impulse-response function which represented the catchment itself. The discharge spectrum was computed for different types of storms, and by using the sampling theorem, the truncation wave number and the network density could be related to each other. In the end, a relationship between the forecasting error and the number of stations is given for each type of storm. This methodology was applied to an Australian basin and the results show that the number of stations required by Eagleson's method was small for reasonable forecasting errors. Eagleson and Goodspeed [1973] used the same method in other Australian basins with similar conclusions.

Bras and Rodríguez-Iturbe [1976] used linear, minimum variance estimation to study the network design problem to estimate an event areal mean. The estimator used was essentially a non-recursive static Kalman filter, which, in turn, ends up being essentially a kriging estimator. Two sources of error make the estimate imprecise: discretization, the spatial

integral of the event is approximated by a finite summation over a regular grid, and interpolation, only a small set of points in the grid are sampled, the remaining points must be estimated by interpolation from the sampled values. The mean square error is then calculated by decomposing it into three terms, i) the model error, that is, the error resulting from approximating a continuous integral by a discrete summation, ii) the estimation error, that is, the error resulting from the interpolation exercise and iii) a term that represents the dependence between the model error and the estimation error. The network design is, then, focused on the minimization of a linear objective function that takes into account network accuracy, station cost and the trade-off between accuracy and cost expressed by the cost that the user is willing to pay for each unit of accuracy. An imaginary case study is presented and several curves of objective function value versus number of stations for different trade-off coefficients are drawn.

To end this review of static network design, the work by Caselton and Husain [1980], who used Shannon's information theory in hydrologic network design, is described. The network is viewed as a communication channel that transmits hydrologic information. The objective of the network design problem is the maximization of information transmission. If X is the random input to the communication channel (the hydrologic event) and Y is the output (the estimate of the hydrologic event), then the information provided by y_j about x_i is related to the change in probability of x_i from the prior value $P[x_i]$ to the posterior value $P[x_i | y_j]$ and the information transmission is measured by the difference between the logarithms of the two probabilities. An expression of the information transmitted by m stations used to interpolate the values at a total of n points was obtained and an exhaustive search was used to devise a daily precipitation network when ten points were available for station location and only some of them were to be used. This method, as well as all the linear estimation methods presented in the preceding paragraphs, have the advantage that the measure of accuracy (in this case the amount of information transmitted) is independent of the measured values themselves, so that the design of the network can be made before any sample is taken. The major disadvantage of this method is the need to estimate all the prior and posterior probability distributions, which

means that the joint probability distribution of input and output to the transmission channel must be known.

All the non-deterministic methods presented in this section claim the advantage that the estimation variance is not dependent on the values of the phenomenon but on its covariance structure and the relative configuration of the sample locations. It is, indeed, an advantage since the network can be designed beforehand; no measurement must be done at the place where a new station is being considered. But, at the same time, it is a disadvantage, since the effectiveness of a station for estimation does not depend only on its location but also on the real measurement. A station can be perfectly allocated with respect to the surrounding stations, but it may end up with a measurement that corresponds to a very improbable or erroneous value that cannot be used in the interpolation process. Some short of data condtioned estimation variance should be used instead.

2.2. Dynamic network design

When time is to be included in the network design problem, so that not only spatial location but sampling frequency is to be determined, then the problem is transformed into one of estimation of dynamic hydrologic processes. The network design problem will then be referred to as dynamic network design.

This part of the literature review is organized in the following way: first, a review of the work done using state estimation theory, and second, a review of the studies that did not use state estimation theory but instead used Monte-Carlo simulations, probability theory, or geostatistics.

State estimation theory provides a framework wherein uncertain dynamic equations can be incorporated into the estimation process. It takes explicit account of error in the model equations and error in the measurements. Only a subset of the state variables are sampled, the remaining are interpolated using the samples and the model predictions. The interpolation function is obtained with the objective of minimizing the mean square error estimate of the

state vector. The Kalman filter is a special case of state estimation theory in which both model equations and measurements are linear in the state variables.

The potential of the Kalman filter in network design was acknowledged in the 70's when it was used by several authors to solve simple cases. But the problem of dimensionality, as reported by Lettenmaier [1979], seems to have doomed its applicability in more recent times. In essence, the Kalman filter consists of two steps: a first step in which the uncertainty in the state variables is propagated into the future by means of the linear model that represents the dynamics of the system, and a second step in which a linear estimation method is used to combine measured and predicted values to obtain the best estimate of the unsampled state variables in the sense of minimum mean square error. Some of the advantages of the Kalman filter are: i) the predicted covariance matrix of the estimation errors is independent of the measurements themselves (as in kriging, this is a direct result from linear estimation theory); ii) measurement errors as well as model errors are accounted for; and iii) the state vector can contain several variables to be sampled in a multidimensional case (cokriging, a special type of kriging, can be used for this purpose, too). On the other hand, dimensionality is a major problem since inversion of very large matrices is involved in the solution.

One of the first researchers to take advantage of the possibilities of the Kalman filter was Moore [1971, 1973]. The objective of his research was "to apply filter theory to the design and improvement of monitoring programs for aquatic ecosystems." the Kalman filter was applied to the measurement of four variables, temperature, zooplankton, phytoplankton and nitrates, in a 12-mile stretch of the American river near Sacramento. Since the dynamics of the variables were not linear, a linearization of the system equations about an operating point was required transforming the method into what is known as an extended Kalman filter. There are two implications of the linearization: the solution obtained will only be suboptimal, and the estimation covariance matrix will depend on the value at the point about which the linearization is being done. Due to lack of experimental data, synthetic data had to be used to evaluate the proposed methodology. The objective was to decide among a set of measurement strategies both in space and time. Two constraints were used, one for the trace of the estimation

variance and another for the cost. The least cost configuration which did not violate any constraint was chosen as the best. As pointed out by Moore himself, computational difficulties and high computer costs were a problem at that time, even though he only looked at 48 state variables (4 variables at 12 locations). In reality both the number of variables to be sampled and the number of locations can be much larger.

Lettenmaier [1975] used the extended Kalman filter for detection of trends in water quality variables. A one-dimensional river model which is steady state in both flow and water quality embodied the dynamics of eight water quality variables (temperature, fecal coliform, nitrate, nitrite, ammonia, orthophosphate, dissolved oxygen and BOD). Only advective terms in the pollutant transport equation were considered. Locations of sampling stations were based on the criterion that the spatial integral over sampling points of the weighted sum (over all the constituents) of the estimation variance be constant. A constraint on the maximum number of samples per unit time was considered. Once the locations of the sampling points were obtained the maximum number of samples per unit time was used to determine the sampling frequency.

Dandy [1976] utilized the Kalman filter in another stream water quality problem. The network design considered the two problems of abatement and prevention, that is, detection of water quality standards violations and determination of long-term changes, respectively. Only BOD and DO were contemplated as state variables. The model of the system was a one-dimensional dynamic flow model incorporating the kinematic wave equation coupled with a transport equation including advective transport (but no dispersive transport) and external and internal sources. Markov processes represented both upstream boundary conditions and tributary and waste stream inflows. A dynamic program was used to search through all the possible design candidates including station location and sampling time. The objective was to maximize the difference between violations detected and cost. Dandy applied this technique to the Shenango-Beaver river in Pennsylvania.

The problem of dimensionality is made worse in the case of groundwater hydrology due to the two- or three-dimensional nature of the problem that rapidly increases the number of

state variables. McLaughlin [1976] used the Kalman filter in estimation and prediction of groundwater levels, although he did not study the network design problem directly. He pointed out the dimensionality problems that lead to computational and numerical difficulties.

The design of a network for runoff prediction was tackled by Bras and Rodríguez-Iturbe [1976]. They suggest a procedure to propagate measurement and interpolation errors from rainfall to runoff through a deterministic rainfall-runoff model. Rainfall intensity was considered as an autoregressive model and was assumed to be observed at a number of discrete points in space. The Kalman filter was then used to obtain the best linear estimation of rainfall intensity and its corresponding estimation covariance matrix. Then, a non-linear state-space model was formulated for the runoff process using a finite difference solution to the kinematic wave equation. After linearization, runoff appears as a linear function of the immediately previous values of runoff and intensity. Using this latter expression a recursive expression for the runoff estimation covariance was derived. This expression was used to evaluate several network configurations. The network configuration directly affected the estimation covariance of rainfall intensity and indirectly the runoff estimation covariance. The methodology was applied to a hypothetical 82-square mile basin schematized by eight overland flow segments and four stream segments. One station could be allocated at each one of the overland segments. Runoff estimation variance was computed as a function of time for several network alternatives.

Kitanidis et al. [1978] shifted from the problem of network design to reduce uncertainty of concentration to the problem of network design to detect violations of water quality standards and proposed two new methodologies, one based on Monte-Carlo simulations and the other based on the Kalman filter. The simulation method consists of the generation of a sufficiently long state history of the system to produce a large statistical sample of violations. Histograms of violations and excursions (adjacent sets of violations) could be calculated at each discrete point in space and a criterion based on these histograms was used to sequentially allocate the sampling points. In the other approach, the Kalman filter formulation was used to compute the asymptotic values, as time tends towards infinity, of the mean and the estimation

covariance of the state variables. Using the asymptotic first and second moments, the asymptotic probability of no violation at a given location at any given time was calculated. Stations should be located sequentially starting with that point with the smallest asymptotic probability of no violation. Once the locations were found using any of the previous approaches, the time interval between samples was computed using the percentage of detected excursions and the percentage of violation points in the detected excursions. The method was used to monitor violations of dissolved oxygen standards in a 54.4 mile stretch of the Shenango-Beaver River, Pennsylvania. The model dynamics were non-linear and a linearization was needed in order to use the Kalman filter formulation.

Pimentel [1975,1978] also focused on asymptotic behavior of the state estimation covariance matrix. Both surface and groundwater pollutant movement were examined for the case that sample spacing is large. Advective movement was not considered. As is pointed out by Lettenmaier [1979], such an assumption cannot be justified except in very particular conditions.

Bras and Colón [1978] used a Kalman filter approach to network design for estimation of long-term areal rainfall average. This problem had been already considered by Rodríguez-Iturbe and Mejía [1974] (see below) using a probabilistic approach. Bras and Colón started from the result proved by Rodríguez-Iturbe and Mejía[1974] that for a stationary process in space and time the long-term space average of rainfall has zero variance and therefore is constant. The rainfall process was then expressed as a function of the residuals about the mean value. Assuming that the residuals follow a multivariate autoregressive model, the dynamics of the rainfall process can be expressed in a convenient way to be included in a discrete Kalman filter formulation (this assumption is equivalent to Rodríguez-Iturbe and Mejía's separable covariance structure). Measurement errors as well as errors in the dynamics of the model were considered. The mean square error matrix was then obtained and it was evaluated for uniform networks with different densities and over different areas. Several autocorrelation parameters of the residuals were tested and trade-off curves between number of stations and intervals of observation for a given value of uncertainty were obtained.

The problem of dynamic network design has also been approached by other authors using other methods, such Grayman and Eagleson [1971], who were interested in network design for runoff prediction using precipitation measurements. The major difference from the work by Eagleson [1967] is the shift from a deterministic rainfall-runoff model to a stochastic one. The main goal of the study was to determine the efficiency of radar in recording rainfall for runoff prediction. A Monte Carlo approach was used that consisted of the following steps: i) generation of rainfall in time and space by means of a stochastic model based on storm characteristics; ii) simulation of the process of recording using radar and radar-rain gage combination; iii) the "recorded" rainfalls values are routed through a rainfall-runoff model based on a spatially distributed solution of the kinematic wave equation. This process was repeated for several sampling schemes focusing on the estimation of peak discharge. Radar errors were divided into large- and small-scale errors and models for both of them were proposed. Point measurement of rainfall was assumed to be fairly accurate. Two of the sampling parameters studied were the required averaging area of the radar signal and the required averaging time. Confidence intervals of predicted discharge for several radar-rain gage combinations were computed. As the number of gages increased the confidence intervals shrunk about the true discharge hydrograph.

Sanders and Adrian [1978] used a probabilistic approach to design a network for detection of water quality trends in rivers. Using standard statistical procedures, the relationship between sampling frequency and the confidence interval of the random component of an annual mean concentration was calculated. River flow was used as a surrogate of water quality. After correlation and non-stationarity have been removed from the data, the remaining residuals can be assumed to be independent and identically distributed. Student's t is used to compute the confidence interval of the theoretical residual mean. The confidence interval of the residual mean is found to be a function of the standard deviation of the observed residuals and of the square root of the number of samples. A plot of the confidence interval versus sampling frequency is proposed to be used to determine the sampling criterion. An example is presented on the Connecticut River near Montague City, Massachusetts.

Bras and Rodríguez-Iturbe [1985, p. 333-358] devote a section of their book to sampling of hydrologic random fields and present a method developed by Rodríguez-Iturbe and Mejía [1974] for the design of precipitation networks. The variance of the long-term mean areal precipitation and mean areal rainfall of a storm event were computed and used to asses the precision associated with the network. For the long-term mean areal rainfall, rainfall is considered as an stationary multidimensional random field with separable covariance function in terms of its spatial and temporal structure. That is, the covariance can be written as a product of three terms, the point covariance, a function of the distance between points, which represents the spatial correlation structure, and a function of the time interval, which represents the temporal correlation structure. The "exact" long-term mean areal rainfall will be an integral over an infinite amount of time and over the whole spatial domain that would be estimated by a finite number of samples distributed in space and time. The variance of this estimate is used as a measure of its precision and it is easily calculable for the special correlation structure considered. The variance of the estimate can finally be expressed as the product of three terms, the point estimation variance, a reduction factor due to the correlation in time, and another reduction factor due to the correlation in space. The time correlation reduction factor depends only on the time interval and not on the sampling location, whereas the space correlation reduction factor depends on space location and not on sampling interval. Three different sampling schemes are considered that lead to three different expressions of the spatial correlation reduction factor: i) random-sampling, in which each station is located randomly with a uniform probability distribution over the whole area, independently of the other stations, ii) stratified random-sampling, in which the whole area is divided into a number of non-overlapping congruent strata and from each stratum k points are randomly selected and iii) systematic sampling, in which the stations are located in a regular geometric pattern. Values of the spatial correlation reduction factor for exponential covariance and for Bessel type covariance illustrated the method. The value of the combined variance reduction factor is studied for several network densities and sampling intervals in a central Venezuela example. Random sampling and Bessel type covariance are considered. The results, for this example,

show that it is more accurate for long-term estimations to take many samples in time over a small number of gages than to take many samples in space over a small time span.

In the same work, Rodríguez-Iturbe and Mejía did a similar analysis of network design for estimation of areal mean of a rainfall event. No time integral is required in this case, so that sampling frequency is not an issue. The approach was exactly the same as the one explained in the previous paragraph only that the time covariance reduction factor did not appear. A most extensive study of possible storms was done.

To end this review, the only application of which I am aware in which geostatistics is used to design a network for estimation of hydrologic dynamic processes will be described. Bastin et al. [1984] proposed a method for optimal estimation of the average areal rainfall and for optimal selection of rain gage locations. Rainfall was considered as a two-dimensional random field with both spatial and temporal variability. The average areal rainfall was computed using kriging, but since time was included in the averaging process, time, as well as space, had to be included in the variogram. Several experimental variograms were computed such as a global time-invariant variogram and monthly variograms. By comparison among them the authors concluded that the variogram model could be separated into a spatial correlation component and a temporal correlation component (very much as Rodríguez-Iturbe and Mejía [1974], did). This special covariance structure makes the estimation variance depend only on the geometry of the sampling location and on a seasonal index. The quotient of estimation variance and the seasonal index turned out to be a time-invariant coefficient that was used to asses the performance of the network. Then, the fictitious point method was applied as in the case of static network design. An example was presented with the objective of reinforcing a set of 17 existing stations in the Semois River basin, Belgium.

3. THE STATE-TRANSITION REPRESENTATION OF A STOCHASTIC LINEAR DYNAMIC SYSTEM

The Kalman filter technique requires a state-transition (also known as state-space) representation of the dynamic system under study. This representation can be obtained for

processes varying in space and time after a spatial discretization of the partial differential equation governing the system has been done. Finite differences or finite elements will yield the following set of ordinary differential equations for processes in which only the first time derivative of the state variables affects the evolution in time of the system (as in the case of flow and mass transport in aquifers),

$$\frac{dx_i(t)}{dt} + a_{1i}(t)x_1(t) + a_{2i}(t)x_2(t) + \cdots + a_{ni}(t)x_n(t) = L_i(t)U_i(t)$$

$$i = 1,...,n \qquad (3.1)$$

where n is the number of discretization points, $x_i(t)$ are the state variables, $U_i(t)$ is a time-varying input or forcing term and $L_i(t)$ is a coefficient. The parameters a_1, a_2,...,a_n may be constant, in with case the result is time invariant. This does not mean that the solution is constant over time, but rather that, given some boundary conditions and some initial conditions at time t_0, the solution at time t_1 does not depend on the values of t_0 and t_1 but only on the difference t_1-t_0.

In the case that the system evolution in time depends on higher-order time derivatives, an extension to the equations developed hereafter is possible and straightforward (see, for instance, Bras and Rodríguez-Iturbe, 1985).

The set of n equations (3.1) can be written in matrix form

$$\dot{\mathbf{x}}(t) = \mathbf{F}(t)\mathbf{x}(t) + \mathbf{L}(t)\mathbf{u}(t) \qquad (3.2)$$

where $\dot{\mathbf{x}}(t)$ is an $n\times1$ vector of the time derivatives of the state variables, $\mathbf{F}(t)$ is an $n\times n$ coefficient matrix and $\mathbf{u}(t)$ represents the external forcing term with dimensions such that the product $\mathbf{L}(t)\mathbf{u}(t)$ is an $n\times1$ vector.

It is recognized that the representation of the system given by (3.1) is not perfect; reasons for this include the spatial discretization of a continuously varying process, uncertainty in the boundary conditions, excessive simplification of reality (which behaves non-linearly), etc. As a result, an error term is introduced in equation (3.2) yielding the canonical form of the state-space representation of a dynamic linear system:

$$\dot{\mathbf{x}}(t) = \mathbf{F}(t)\mathbf{x}(t) + \mathbf{L}(t)\mathbf{u}(t) + \mathbf{G}(t)\mathbf{w}(t) \tag{3.3}$$

where $\mathbf{w}(t)$ is white noise, $\mathbf{G}(t)$ is such that $\mathbf{G}(t)\mathbf{w}(t)$ has dimension $n\times 1$, and mean and covariance given by

$$E\{\mathbf{G}(t)\mathbf{w}(t)\} = \mathbf{0}$$

$$E\{(\mathbf{G}(t)\mathbf{w}(t))(\mathbf{G}(\tau)\mathbf{w}(\tau))^T\} = \mathbf{G}(t)\mathbf{Q}(t)\mathbf{G}(t)^T\delta(t-\tau)$$

where $\delta(t-\tau)$ is the Dirac delta function, $\mathbf{Q}(t)$ is called a spectral-density matrix, which becomes a covariance matrix when multiplied by the Dirac delta function. I will not enter here into a discussion of white noise processes (see, for instance, Hoel et al., 1972, section 5.4, or Schweepe, 1973, Appendix F), I just want to note that a continuous time stochastic white noise process does not have a physical sense, since it is a process with zero mean and infinite covariance matrix. A continuous stochastic white noise can be interpreted as the time derivative of a stochastic process with uncorrelated increments (i.e, a Wiener process).

The discrete state-transition model is obtained by integration of (3.3) over a time interval Δt

$$\mathbf{x}(t+\Delta t) = \mathbf{\Phi}(t+\Delta t, t)\mathbf{x}(t) + \mathbf{\Lambda}(t)\mathbf{u}(t) + \mathbf{\Gamma}(t)\mathbf{w}(t) \tag{3.4}$$

where the matrices $\mathbf{\Phi}$, $\mathbf{\Lambda}$, and $\mathbf{\Gamma}$ are related to matrices $\mathbf{F}$, $\mathbf{L}$ and $\mathbf{G}$ by (see, for example, Bras and Rodríguez-Iturbe, 1985)

$$\Phi(t, t+\Delta t) = e^{\mathbf{F}\cdot\Delta t}$$

$$\Lambda(t)\mathbf{u}(t) = \int_t^{t+\Delta t} \Phi(t,\tau)\mathbf{L}(\tau)\mathbf{u}(\tau)\,d\tau$$

$$\Gamma(t)\mathbf{w}(t) = \int_t^{t+\Delta t} \Phi(t,\tau)\mathbf{G}(\tau)\mathbf{w}(\tau)\,d\tau$$

The last integral does not have a Riemann interpretation due to the stochastic nature of the white process $\mathbf{w}(t)$. Itô calculus is necessary (see, for instance, Hoel et al. 1972). Equation (3.4) is generally rewritten as

$$\mathbf{x}(k+1) = \Phi(k)\mathbf{x}(k) + \Lambda(k)\mathbf{u}(k) + \Gamma(k)\mathbf{w}(k) \qquad (3.5)$$

with $t=k\cdot\Delta t$ and understanding that $\Phi(k)$ refers always to the transition over one time step. The vector $\mathbf{w}(k)$, $k=1,...,n$ is called a discrete time white process, which is a process with mutually independent random vectors, with zero mean and covariance

$$E\{\mathbf{w}(k_1)\mathbf{w}^T(k_2)\} = \begin{cases} \mathbf{Q}(k_1), & k_1 = k_2 \\ \mathbf{0}, & k_1 \neq k_2 \end{cases}$$

In the rest of this chapter the equation (3.5) will be derived for groundwater flow and mass transport.

3.1. State-transition model for groundwater flow

An equation equivalent to (3.5) will be obtained by discretization and further manipulation of the classical equation for vertically-averaged 2-dimensional horizontal flow in a vertically homogeneous, anisotropic aquifer with principal axes of the transmissivity in the x, y directions

$$\frac{\partial}{\partial x}\left(T_x \frac{\partial h}{\partial x}\right) + \frac{\partial}{\partial y}\left(T_y \frac{\partial h}{\partial y}\right) = S \frac{\partial h}{\partial t} + Q \qquad (3.6)$$

where x and y are the spatial coordinates (L), t is time (T), h is the piezometric head in the aquifer (L), T_x and T_y are transmissivities (L/T^2) (respectively equal to the integral of the hydraulic conductivity K_x or K_y over the saturated thickness), S is the specific storage (L^3/L^3) and Q is the net external discharge per unit area ($L^3/T/L^2$).

Finite differences or finite elements can be used (see Remson et al., 1971; Trescott et al., 1980 or Lapidus and Pinder, 1984) to spatially discretize equation (3.6) over a number of nodes or elements. The result is a set of simultaneous algebraic equations of the form

$$\mathbf{\Phi}'' \mathbf{h} + \mathbf{\vartheta}'' \frac{d\mathbf{h}}{dt} = \mathbf{\Lambda}'' \mathbf{u} \qquad (3.7)$$

where $\mathbf{h}$ is an $n\times1$ vector, n being the number of internal nodes in the aquifer, $\mathbf{\Phi}''$ is an $n\times n$ matrix function of the transmissivity values and the discretization (this matrix is also function of time in the case of an unconfined aquifer, or time-dependent head boundaries), $\mathbf{\vartheta}''$ is an $n\times n$ matrix function of the storage, $\mathbf{\Lambda}''$ is a coefficient matrix of arbitrary dimensions and $\mathbf{u}$ is a vector with the net discharges at each node and the specified fluxes at the Neumann boundary nodes. The dimensions of $\mathbf{\Lambda}''$ and $\mathbf{u}$ must be such that the product is an $n\times1$ vector. Multiplying by the left by the inverse of $\mathbf{\vartheta}''$ (a diagonal matrix in the case of finite differences) a new matrix equation results

$$\frac{d\mathbf{h}}{dt} = \mathbf{\Phi}'\mathbf{h} + \mathbf{\Lambda}'\mathbf{u} \qquad (3.8)$$

where now both $\mathbf{\Phi}'$ and $\mathbf{\Lambda}'$ are also functions of the storage. In the case of second order central finite differences over a uniform rectangular grid, the matrix $\mathbf{\Phi}'$ is defined by the following algorithm

For each node i there will be the following nonzero elements in row i

• in column i

$$\Phi'_{i,i} = \frac{-1}{S_i}\left(\frac{T_{x,\,i+1/2} + T_{x,\,i-1/2}}{\Delta x^2} + \frac{T_{y,\,i+1/2} + T_{y,\,i-1/2}}{\Delta y^2} \right)$$

• in the column corresponding to the nearest cell in the positive x direction, (in general, cell i+1)

$$\Phi'_{i,\,i+1} = \frac{T_{x,\,i+1/2}}{S_i\ \Delta x^2}$$

• in the column corresponding to the nearest cell in the negative x direction (in general, cell i-1)

$$\Phi'_{i,\,i-1} = \frac{T_{x,\,i-1/2}}{S_i\ \Delta x^2}$$

• in the column corresponding to the nearest cell in the positive y direction (in general, cell i+k, k being the number of grid columns)

$$\Phi'_{i,\,i+k} = \frac{T_{y,\,i+1/2}}{S_i\ \Delta y^2}$$

• in the column corresponding to the nearest cell in the negative y direction (in general, cell i -k)

$$\Phi'_{i,\,i-k} = \frac{T_{y,\,i-1/2}}{S_i\ \Delta y^2}$$

where, Δx and Δy are the cell dimensions in the x and y directions, S_i is the storage coefficient at node i, $T_{x,\,i+1/2}$ is the transmissivity between node i and the nearest node in the positive x

direction, $T_{x,i\text{-}1/2}$ is the transmissivity between node i and the nearest node in the negative x direction and similarly for the transmissivities in the y direction.

Before proceeding, a number of assumptions and their consequences must be stated: i)T_x , T_y, S and Q, as well as the specified flux of the Neumann boundary conditions, are perfectly known, ii) the model is imperfect and such imperfection will be represented by an additive discrete white process, iii) the initial conditions are known with uncertainty given by an estimation covariance matrix at time zero, iv) the uncertainty in initial heads may extend to the constant head boundaries. Assumption i) means that $\mathbf{\Phi}'$ and $\mathbf{\Lambda}'$ are deterministic, known coefficient matrices, and that $\mathbf{u}$ represents deterministic known net discharges. Assumption ii) introduces the term $\mathbf{\Gamma}'\mathbf{w}'$ in the equation to represent model error. Assumption iii) introduces uncertainty in the heads from time zero, and assumption iv) allows the inclusion of uncertain constant head boundaries. The introduction of model error in equation (3.8) yields

$$\frac{d\,\mathbf{h}}{dt} = \mathbf{\Phi}'\mathbf{h} + \mathbf{\Lambda}'\mathbf{u} + \mathbf{\Gamma}'\mathbf{w}'$$

The discretization of the time derivative will finally lead to the desired state-transition equation

$$\mathbf{h}(k+1) = \mathbf{\Phi}(k)\mathbf{h}(k) + \mathbf{\Lambda}(k)\mathbf{u}(k) + \mathbf{\Gamma}(k)\mathbf{w}(k) \qquad (3.9)$$

with $t=k\Delta t$. The final form of the state-transition matrix as well as of the coefficient matrices will depend on the type of time discretization chosen. If an explicit approximation is preferred

$$\frac{d\,\mathbf{h}(k)}{dt} \approx \frac{\mathbf{h}(k+1) - \mathbf{h}(k)}{\Delta t}$$

then

$$\Phi(k) = [\Delta t \cdot \Phi'(k) + \mathbf{I}]$$
$$\Lambda(k) = [\Delta t \cdot \Lambda'(k)] \qquad (3.10)$$
$$\Gamma(k) = [\Delta t \cdot \Gamma'(k)]$$

where **I** is the identity matrix. If the fully implicit method is chosen

$$\frac{d\,\mathbf{h}(k)}{dt} \approx \frac{\mathbf{h}(k) - \mathbf{h}(k-1)}{\Delta t}$$

then

$$\Phi(k) = [\mathbf{I} - \Delta t \cdot \Phi'(k)]^{-1}$$
$$\Lambda(k) = \Phi(k)\,[\Delta t \cdot \Lambda'(k)] \qquad (3.11)$$
$$\Gamma(k) = \Phi(k)\,[\Delta t \cdot \Gamma'(k)]$$

3.2. State-transition model for mass transport

Following the same steps as for the flow case, the equation for vertically integrated 2-dimensional mass transport will be manipulated until a state transition equation is obtained. The equation used is

$$\frac{\partial}{\partial x}\left(D_{xx}\frac{\partial c}{\partial x} + D_{xy}\frac{\partial c}{\partial y}\right) + \frac{\partial}{\partial y}\left(D_{xy}\frac{\partial c}{\partial x} + D_{yy}\frac{\partial c}{\partial y}\right)$$
$$- \frac{\partial}{\partial x}\left(v_x c\right) - \frac{\partial}{\partial y}\left(v_y c\right) = \frac{\partial c}{\partial t} + \frac{W}{\phi} \qquad (3.12)$$

where x and y are the spatial coordinates (L), t is time (T), c is the vertical integral of the concentration (ML^{-2}), v_x and v_y are the components of the fluid velocity in the x and y

directions (L /T), W is the net external efflux of contaminant concentration per unit area (M /L^2T), ϕ is the porosity (L^3/L^3), D_{xx} , D_{xy} and D_{yy} are the components of the dispersion tensor

$$\mathbf{D} = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix}$$

(L^2/T). The dispersion coefficients are functions of the fourth order dispersivity tensor (a_{ijkm}) and of the velocity. In the case of on isotropic medium the dispersivity tensor is expressed in terms of two constants, a_L the longitudinal dispersivity, and a_T the transversal dispersivity (see, Bear, 1979) and in that case the dispersion coefficients are given by

$$\begin{aligned} D_{xx} &= (a_L v_x^2 + a_T v_y^2)/v \\ D_{xy} &= (a_L - a_T) v_x v_y /v \\ D_{yy} &= (a_L v_y^2 + a_T v_x^2)/v \\ v &= \sqrt{v_x^2 + v_y^2} \end{aligned} \qquad (3.13)$$

A finite difference or a finite element scheme can be used to spatially discretize equation (3.12) (see Pinder and Gray, 1977 or Lapidus and Pinder, 1982). The velocity field is computed using Darcy's law from the results of a previous solution of a groundwater flow model.

$$v_{i,i+1} = - \frac{K_{i,i+1}}{\phi} \left(\frac{h_{i+1} - h_i}{\Delta(i)} \right)$$

where, given two adjacent nodes i and i+1, $v_{i,i+1}$ is the flow velocity from node i to node i+1, $K_{i,i+1}$ is the hydraulic conductivity between nodes, ϕ is a representative value of the porosity along the path between both nodes (the average value of the porosities at both nodes is used),

h_i and h_{i+1} are the respective piezometric heads, and Δ (i) is the distance between nodes. A major problem arises when central differences are used to discretize equation (3.12) since the components of **D** must be approximated at the center point between nodes and to compute them an approximation of the velocity vector is needed. Using Darcy's law and the head results from a previous groundwater model, it is possible to obtain the component of the velocity in the direction from node to node and assign it to the center point between nodes, but the velocity may have a nonzero orthogonal component. It is customary to assume that that unknown component can be obtained by bilinear interpolation of the four nearest between-node velocities in its same direction. A similar problem happens with the concentrations, since they also need to be known at the center point between nodes; a linear interpolation of the two nearest concentration values is used. With all these interpolations in mind, and careful algebraic manipulations, an equation similar to (3.8) can be obtained for concentrations

$$\frac{d\,\mathbf{c}}{dt} = \mathbf{\Phi}' \mathbf{c} + \mathbf{\Lambda}' \mathbf{u} \qquad (3.14)$$

where $\mathbf{\Phi}'$ is defined by the following algorithm for the case of finite differences and rectangular grid with constant column and row widths.

- Each node in the grid is referred by its row i and column j .
- Row numbers grow in the positive y direction and column numbers grow in the positive x direction.
- There are l rows and m columns for a total of $n=l \times m$ nodes
- The node (i,j) corresponds to position k in the transfer matrix,
  $k = (i\text{-}1)\times m + j$

$$\mathbf{\Phi}'_{k,k-m-1} = \frac{1}{4\Delta x\,\Delta y}\left(D_{xy,\,ij\,-1/2} + D_{xy,\,i\,-1/2j}\right)$$

$$\Phi'_{k,k-m} = \frac{1}{4\Delta x\,\Delta y}\left(-D_{xy,\,ij+1/2} + D_{xy,\,ij-1/2}\right) + \frac{D_{yy,\,i-1/2j}}{(\Delta y)^2} + \frac{v_{x,\,i-1/2j}}{2\Delta y}$$

$$\Phi'_{k,k-m+1} = \frac{1}{4\Delta x\,\Delta y}\left(-D_{xy,\,ij+1/2} - D_{xy,\,i-1/2j}\right)$$

$$\Phi'_{k,k-1} = \frac{1}{4\Delta x\,\Delta y}\left(-D_{xy,\,i+1/2j} + D_{xy,\,i-1/2j}\right) + \frac{D_{xx,\,ij-1/2}}{(\Delta x)^2} + \frac{v_{x,\,ij-1/2}}{2\Delta x}$$

$$\Phi'_{k,k} = \frac{1}{(\Delta x)^2}\left(-D_{xx,\,ij+1/2} - D_{xx,\,ij-1/2}\right) + \frac{1}{(\Delta y)^2}\left(-D_{yy,\,i+1/2j} - D_{yy,\,i-1/2j}\right)$$

$$+ \frac{1}{2\Delta x}\left(-v_{x,\,ij+1/2} + v_{x,\,ij-1/2}\right) + \frac{1}{2\Delta y}\left(-v_{y,\,i+1/2j} + v_{y,\,i-1/2j}\right)$$

$$\Phi'_{k,k+1} = \frac{1}{4\Delta x\,\Delta y}\left(D_{xy,\,i+1/2j} - D_{xy,\,i-1/2j}\right) + \frac{D_{xx,\,ij+1/2}}{(\Delta x)^2} - \frac{v_{y,\,ij+1/2}}{2\Delta x}$$

$$\Phi'_{k,k-m+1} = \frac{1}{4\Delta x\,\Delta y}\left(-D_{xy,\,ij-1/2} - D_{xy,\,i+1/2j}\right)$$

$$\Phi'_{k,k+m} = \frac{1}{4\Delta x\,\Delta y}\left(-D_{xy,\,ij-1/2} + D_{xy,\,ij+1/2}\right) + \frac{D_{yy,\,i+1/2j}}{(\Delta y)^2} - \frac{v_{y,\,i+1/2j}}{2\Delta y}$$

$$\Phi'_{k,k+m+1} = \frac{1}{4\Delta x\,\Delta y}\left(D_{xy,\,ij+1/2} + D_{xy,\,i+1/2j}\right)$$

where the dispersion coefficients are defined according to (3.13) with the following velocities

$$v_{x,\,ij+1/2} = \frac{K_{xx,\,i,j+1/2}}{\phi}\left(\frac{h_{ij} - h_{ij+1}}{\Delta x}\right)$$

$$v_{x,\,ij-1/2} = \frac{K_{xx,\,i,j-1/2}}{\phi}\left(\frac{h_{ij-1} - h_{ij}}{\Delta x}\right)$$

$$v_{y,\,i+1/2j} = \frac{K_{yy,\,i+1/2j}}{\phi}\left(\frac{h_{ij} - h_{i+1j}}{\Delta y}\right)$$

$$v_{y,\,i-1/2j} = \frac{K_{yy,\,i-1/2j}}{\phi}\left(\frac{h_{i-1j} - h_{ij}}{\Delta y}\right)$$

$$v_{y,ij+1/2} = \frac{1}{4}\left(v_{y,i-1/2j} + v_{y,i-1/2j+1} + v_{y,i+1/2j} + v_{y,i+1/2j+1}\right)$$

$$v_{y,ij-1/2} = \frac{1}{4}\left(v_{y,i-1/2j} + v_{y,i-1/2j-1} + v_{y,i+1/2j} + v_{y,i+1/2j-1}\right)$$

$$v_{x,i+1/2j} = \frac{1}{4}\left(v_{y,ij-1/2} + v_{y,i+1j-1/2} + v_{y,ij+1/2} + v_{y,i+1j+1/2}\right)$$

$$v_{x,i-1/2j} = \frac{1}{4}\left(v_{y,ij-1/2} + v_{y,i-1j-1/2} + v_{y,ij+1/2} + v_{y,i-1j+1/2}\right)$$

The same sort of assumptions that were invoked for the flow model are invoked here, namely, perfect knowledge of all parameter values (including velocities) as well as of any specified mass fluxes at boundaries, uncertainty in initial conditions, including constant concentration boundaries, and the possibility of model error. Addition of a discrete white error to equation (3.14) results in

$$\frac{d\mathbf{c}}{dt} = \Phi'\mathbf{c} + \Lambda'\mathbf{u} + \Gamma'\mathbf{w}'$$

Discretization of the time derivative of concentration will lead to the state-transition equation

$$\mathbf{c}(k+1) = \Phi(k)\mathbf{c}(k) + \Lambda(k)\mathbf{u}(k) + \Gamma(k)\mathbf{w}(k) \qquad (3.15)$$

with the same relationship between Φ, Λ and Γ and Φ', Λ' and Γ' as in equations (3.10) and (3.11), depending on the time discretization scheme used.

4. The Kalman filter

The Kalman filter, also known as the Kalman-Bucy filter, was developed in the early sixties by Kalman [1960] and Kalman and Bucy [1961] with the objective of obtaining an optimal estimate, for a given loss function, of a random function that evolves in time according to a linear dynamic system. Model errors as well as measurement errors were also considered. The possibility of symmetric loss functions other than the square of the estimation error were

discussed in the paper by Kalman [1960], and there he pointed out that the optimal estimate obtained using the square of the estimation error is valid for several other symmetric loss functions. The square of the estimation error is most often used and is the one considered in this work.

The basic idea in the development of the Kalman filter theory was to combine the state-space representation of a stochastic linear dynamic system and linear estimation regarded as an orthogonal projection in Hilbert space. The details of the Kalman filter will be presented later using a different approach than the one used by Kalman [1960].

The work by Kalman soon found application in a variety of fields, ranging from circuits to ballistics to groundwater hydrology, just to give a few examples. Extensions to the original work have also been made to deal with non-linear systems or to include parameter uncertainty in the filtering process. Non-linear filters could be the next extension, but the difficulty and cost of non-linear filters used in static systems (such as disjunctive kriging or non-parametric geostatistics) might prevent them from being used with dynamic systems. All extensions rely on the original contribution by Kalman and Bucy of coupling the state-space representation with optimal estimation theory to approach dynamic systems.

4.1. Estimation of static systems.

The development of the Kalman equations in the original paper made extensive use of projection theory and Hilbert space. Here I will use the same approach as Schweppe [1973] and Bras and Rodríguez-Iturbe [1985]. This approach requires first the development of the optimum linear estimate of a static system from a set of measurements. The variables to be estimated are represented by the *n* dimensional state vector **x**, the observations by the *m* dimensional vector **z** and a mathematical model relates observations to measurements. The mathematical model consists of an observation equation and an uncertainty model:

1. Observation equation

   $\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{v}$

   x is an *n* dimensional vector with the true, but unknown values of the system,

v is a *m* vector of measurement errors which are also unknown,

H is a known $m \times n$ matrix called the observation matrix, and

z is an *m* dimensional vector with the actual measurements.

2. Two models will be considered for the uncertainty in **x** and **v**:

   a. Bayesian model. **v** is a random vector with known mean and variance, and **x** is a random vector for which a priori information is available, in particular its mean and variance.

   b. Fisher model. **v** is a random vector with known mean and variance, and **x** is a constant (not a random variable) that may vary in space and for which no a priori information is available.

Since the true value of **x** is unknown, an estimate $\hat{\mathbf{x}}$ (the hat will be used to denote estimation) is to be found as a function of the observations and the mathematical model (namely, observation matrix **H** and uncertainty models for **x** and **v**). The usual way to approach this problem is to try to express the estimate $\hat{\mathbf{x}}$ as a function of the measurements **z**. The function will be determined by the mathematical model. In this work only linear estimates are used, so the estimation can be expressed in matrix form as:

$\hat{\mathbf{x}} = \mathbf{W}\mathbf{z} + \mathbf{W_0}$

W is an $n \times m$ matrix determined by the mathematical model and

$\mathbf{W_0}$ is an $n \times 1$ vector of coefficients.

The criteria to determine **W** and $\mathbf{W_0}$ can be many but it is reasonable to look for **W** and $\mathbf{W_0}$ such that some loss function is minimized. Loss functions are in general a function of the estimation error $\hat{\mathbf{x}}$-**x**. Here, the mean square estimation error will be used. Thus **W** and $\mathbf{W_0}$ will be computed in such a way that it minimizes the expected value of the square of the estimation error. It has been proven elsewhere, see for instance Papoulis [1984], that the function, linear or non-linear, that provides the best estimate of **x** according to the previous criterion is the mean of the conditional distribution of **x** given the set of observations. It has also been proven

that the linear estimation will yield the best estimation possible from the set of observations when **x** and **v** are jointly normally distributed (Papoulis, 1984).

4.1.1. Bayesian model

In the bayesian model **x** is assumed to be a random variable for which some a priori information is available. The uncertainty in the random variables **x** and **v** is characterized by the first two moments of their joint distribution,

$$\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{v} \quad \text{(Observation eq.)}$$

$$E\{\mathbf{x}\} = \mathbf{m}$$

$$E\{(\mathbf{x} - \mathbf{m})(\mathbf{x} - \mathbf{m})^T\} = \mathbf{\Psi}$$

$$E\{\mathbf{v}\} = \mathbf{0}$$

$$E\{\mathbf{v}\mathbf{v}^T\} = \mathbf{R}$$

$$E\{\mathbf{x}\mathbf{v}^T\} = \mathbf{0}$$

Hence, the measurement error is represented by a white noise that is uncorrelated with the state variable. The mean and the covariance matrices of both **x** and **v** are known.

For this model, the unbiased linear estimator $\hat{\mathbf{x}}$ that minimizes the error covariance matrix is given by a linear function of the form

$$\hat{\mathbf{x}} = \mathbf{W}\mathbf{z} + \mathbf{W}_0 \qquad (4.1)$$

where **W** and $\mathbf{W}_0$ are given by

$$W = \Sigma H^T R^{-1}$$
$$W_0 = \Sigma \Psi^{-1} m$$
$$\Sigma = (H^T R^{-1} H + \Psi^{-1})^{-1}$$

It can be shown that the unbiased condition is met

$$E\{\hat{x}\} = m$$

and that Σ is the minimum error covariance matrix,

$$\Sigma = E\{(x - \hat{x})(x - \hat{x})^T\}$$

The error covariance matrix is minimum in the sense that if any other set $\tilde{W}$, $\tilde{W}_0$ is used to obtain another estimate

$$\tilde{x} = \tilde{W} z + \tilde{W}_0$$

the error covariance matrix associated with this new estimate

$$\tilde{\Sigma} = E\{(x - \tilde{x})(x - \tilde{x})^T\}$$

is larger than Σ in the sense that $\tilde{\Sigma} - \Sigma$ is non-negative definite [Golub,1983]. A detailed proof of this result can be found in Liebelt [1967, ch.5].

After some matrix manipulation (described in detail in Liebelt[1967]), the optimal estimator can be rewritten as

$$\hat{x} = m + \Psi H^T [H \Psi H^T + R]^{-1} [z - Hm]$$
$$\Sigma = E\{(x - \hat{x})(x - \hat{x})^T\}$$
$$= \Psi - \Psi H^T [H \Psi H^T + R]^{-1} H \Psi$$

These same results are found in many other fields with, apparently, different interpretations. To cite an example, linear geostatistics is another way to approach estimation of static systems and simple kriging ends up with exactly these same equations when the measurements are assumed noiseless.

4.1.2. Fisher model

In the Fisher model, **x** is assumed to be a constant, not a random variable (although this constant may vary from one point to another in space). The observation equation remains the same but now there is no a priori information about the state variables,

$$\begin{aligned} \mathbf{z} &= \mathbf{H}\mathbf{x} + \mathbf{v} \\ &\mathbf{x} \text{ is completely unknown} \\ E\{\mathbf{v}\} &= \mathbf{0} \\ E\{\mathbf{v}\mathbf{v}^T\} &= \mathbf{R} \end{aligned}$$

The measurement error is, again, a white noise of known covariance matrix.

The unbiased linear estimation $\hat{\mathbf{x}}$ that minimizes the error covariance matrix is given by the function

$$\hat{\mathbf{x}} = \mathbf{W}\mathbf{z} + \mathbf{W}_0 \tag{4.2}$$

where **W**, $\mathbf{W}_0$ are given by

$$\begin{aligned} \mathbf{W} &= \boldsymbol{\Sigma}\, \mathbf{H}^T\mathbf{R}^{-1} \\ \mathbf{W}_0 &= \mathbf{0} \end{aligned}$$

if $\boldsymbol{\Sigma}$ exists,

$$\Sigma = E\{(\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^T\}$$

$$= (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}$$

The necessary condition for $\boldsymbol{\Sigma}$to exist is that the dimension *m* of the observation vector **z** must not be smaller than the dimension *n* of the state vector **x** (see Liebelt[1967, ch. 5] for detailed proof). In this case the estimation is unbiased in the sense that,

$$E\{\hat{\mathbf{x}}\} = \mathbf{x}$$

4.1.3. Two observations of the same state

Before proceeding to the linear estimation of a dynamic system, let us consider the case of estimation of a static system for which two sets of observations are available. Again, the cases of Bayesian and Fisher estimation will be considered.

Assume that two observations of the same state vector are available

$$\mathbf{z}_1 = \mathbf{H}_1\mathbf{x} + \mathbf{v}_1 \tag{4.3}$$

$$\mathbf{z}_2 = \mathbf{H}_2\mathbf{x} + \mathbf{v}_2 \tag{4.4}$$

These two observations can be viewed as a single observation by augmenting the observation vector

$$\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{v}$$

where

$$\mathbf{z} = \begin{bmatrix} \mathbf{z}_1 \\ \dots \\ \mathbf{z}_2 \end{bmatrix} \qquad \mathbf{H} = \begin{bmatrix} \mathbf{H}_1 \\ \dots \\ \mathbf{H}_2 \end{bmatrix} \qquad \mathbf{v} = \begin{bmatrix} \mathbf{v}_1 \\ \dots \\ \mathbf{v}_2 \end{bmatrix}$$

and proceed as a single vector observation. However, it may be interesting to regard the problem as a two-step process: first, obtain separate estimates of $\mathbf{x}$, one from $\mathbf{z_1}$ and the other from $\mathbf{z_2}$, and second, combine those estimates to obtain the final one. It will be shown how these two estimates must be made and how they must be combined to yield the same estimate as if the two observations were considered as a single one. Both the Fisher and the Bayesian model will be contemplated.

First, consider the Fisher model where the state $\mathbf{x}$ is an unknown constant and the observation errors are zero mean, uncorrelated random vectors,

$$E\{\mathbf{v}_1\mathbf{v}_1^T\}=\mathbf{R}_1$$

$$E\{\mathbf{v}_1\mathbf{v}_2^T\}=\mathbf{0}$$

$$E\{\mathbf{v}_2\mathbf{v}_2^T\}=\mathbf{R}_2$$

If $\mathbf{z}_1$ and $\mathbf{z}_2$ are combined in a single, augmented vector and equation (4.2) is used, the estimate and its estimation error are,

$$\begin{aligned}\hat{\mathbf{x}} &= \Sigma\,[\mathbf{H}_1^T\mathbf{R}_1^{-1}\mathbf{z}_1+\mathbf{H}_2^T\mathbf{R}_2^{-1}\mathbf{z}_2]\\ \Sigma &= [\mathbf{H}_1^T\mathbf{R}_1^{-1}\mathbf{H}_1+\mathbf{H}_2^T\mathbf{R}_2^{-1}\mathbf{H}_2]^{-1}\end{aligned} \qquad (4.5)$$

Consider now the Fisher estimates of $\mathbf{x}$ using $\mathbf{z}_1$ and $\mathbf{z}_2$ separately. These will be:

$$\hat{\mathbf{x}}_1=\Sigma_1\mathbf{H}_1^T\mathbf{R}_1^{-1}\mathbf{z}_1$$
$$\Sigma_1=[\mathbf{H}_1^T\mathbf{R}_1^{-1}\mathbf{H}_1]^{-1}$$

and

$$\hat{\mathbf{x}}_2=\Sigma_2\mathbf{H}_2^T\mathbf{R}_2^{-1}\mathbf{z}_2$$
$$\Sigma_2=[\mathbf{H}_2^T\mathbf{R}_2^{-1}\mathbf{H}_2]^{-1}$$

These two estimates can be combined, to give the same result as (4.5), as follows,

$$\hat{\mathbf{x}} = [\Sigma_1^{-1} + \Sigma_2^{-1}]^{-1} [\Sigma_1^{-1} \hat{\mathbf{x}}_1 + \Sigma_2^{-1} \hat{\mathbf{x}}_2]$$
$$\Sigma = [\Sigma_1^{-1} + \Sigma_2^{-1}]^{-1}$$

Notice, that this estimate corresponds to the Fisher estimate of **x** using equation (4.5) when two observations $\hat{\mathbf{x}}_1$ and $\hat{\mathbf{x}}_2$ (instead of $\mathbf{z}_1$ and $\mathbf{z}_2$) are made according to the observation equations

$$\hat{\mathbf{x}}_1 = \mathbf{x} + \delta \mathbf{x}_1$$

$$\hat{\mathbf{x}}_2 = \mathbf{x} + \delta \mathbf{x}_2$$

where the matrices $\mathbf{H}_1$ and $\mathbf{H}_2$ are the identity matrix and the errors are $\delta \mathbf{x}_1$ and $\delta \mathbf{x}_2$ with covariance matrices

$$E\{\delta \mathbf{x}_1 \delta \mathbf{x}_1^T\} = \Sigma_1$$

$$E\{\delta \mathbf{x}_2 \delta \mathbf{x}_2^T\} = \Sigma_2$$

Therefore, the Fisher estimate of **x** can be obtained in either of two ways, 1) by combining the two observations (4.3) and (4.4) in one and then using equation (4.2), or 2) by the two-step process consisting in, first obtaining the Fisher estimates corresponding to each observation separately and second obtaining the Fisher estimate that would correspond to these two last estimates regarded as two observations of the system (see Figure 4.1).

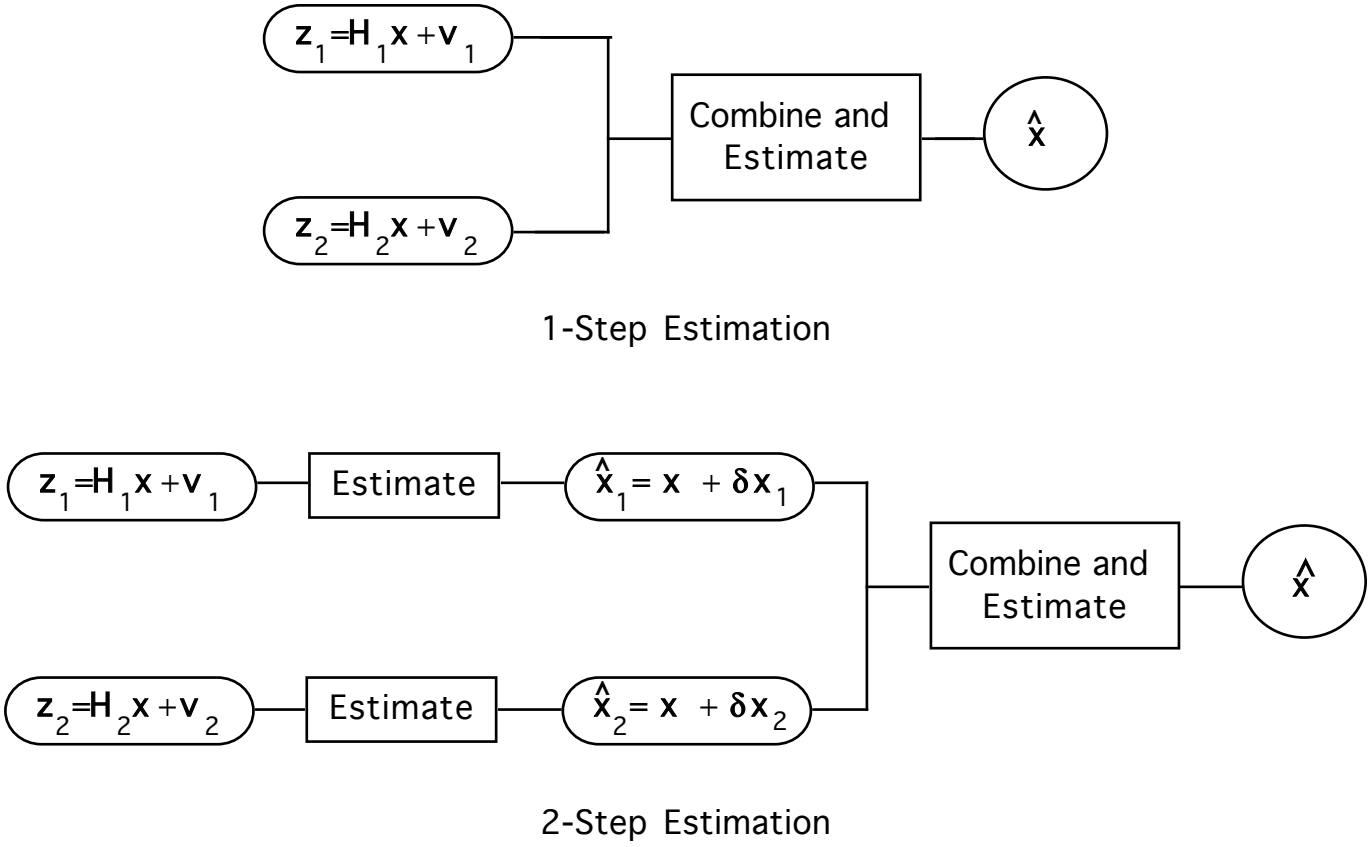


Figure 4.1. Estimation from two different observations.

Now consider the Bayesian model where **x** is a random vector with known mean and covariance and uncorrelated with $\mathbf{v}_1$ and $\mathbf{v}_2$,

$$E\{\mathbf{x}\} = \mathbf{m}$$
$$E\{(\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^T\} = \Psi$$

The combination of the two observations in one and the use of equation (4.1) yields the estimate:

$$\begin{aligned} \hat{\mathbf{x}} &= \Sigma\,[\mathbf{H}_1^T\mathbf{R}_1^{-1}\mathbf{z}_1 + \mathbf{H}_2^T\mathbf{R}_2^{-1}\mathbf{z}_2 + \Psi^{-1}\mathbf{m}] \\ \Sigma &= [\mathbf{H}_1^T\mathbf{R}_1^{-1}\mathbf{H}_1 + \mathbf{H}_2^T\mathbf{R}_2^{-1}\mathbf{H}_2 + \Psi^{-1}]^{-1} \end{aligned} \tag{4.6}$$

The kind of two-step estimation shown in Figure 4.1 still can be made if we assume that the a priori information (mean and covariance of **x**) is associated with one of the observations, say

$\mathbf{z}_1$, while the second observation is viewed as if $\mathbf{x}$ were a constant completely unknown rather than a random variable.

Then, the Bayesian estimate associated with observation $\mathbf{z}_1$ using equation (4.1) would be,

$$\hat{\mathbf{x}}_1 = \Sigma_1 [H_1^T R_1^{-1} Z_1 + \Psi^{-1} m]$$
$$\Sigma_1 = [H_1^T R_1^{-1} H_1 + \Psi^{-1}]^{-1}$$

whereas the Fisher estimate associated with observation $\mathbf{z}_2$ using equation (4.2) would be,

$$\hat{\mathbf{x}}_2 = \Sigma_2 [H_2^T R_2^{-1} z_2]$$
$$\Sigma_2 = [H_2^T R_2^{-1} H_2]^{-1}$$

Now equation (4.6) can be rewritten in terms of these two estimates as

$$\hat{\mathbf{x}} = [\Sigma_1^{-1} + \Sigma_2^{-1}]^{-1} [\Sigma_1^{-1} \hat{\mathbf{x}}_1 + \Sigma_2^{-1} \hat{\mathbf{x}}_2]$$
$$\Sigma = [\Sigma_1^{-1} + \Sigma_2^{-1}]^{-1}$$

which again corresponds to the Fisher estimate of $\mathbf{x}$ using equation (4.5) when two observations $\hat{\mathbf{x}}_1$ and $\hat{\mathbf{x}}_2$ (instead of $\mathbf{z}_1$ and $\mathbf{z}_2$) are made according to the observation equations

$$\hat{\mathbf{x}}_1 = \mathbf{x} + \delta \mathbf{x}_1$$

$$\hat{\mathbf{x}}_2 = \mathbf{x} + \delta \mathbf{x}_2$$

with observation errors $\delta \mathbf{x}_1$ and $\delta \mathbf{x}_2$ with covariance matrices

$$E\{\delta \mathbf{x}_1 \delta \mathbf{x}_1^T\} = \Sigma_1$$

$$E\{\delta \mathbf{x}_2 \delta \mathbf{x}_2^T\} = \Sigma_2$$

Therefore, the Bayesian estimate of **x** when two measurements are available, can be obtained in either of two ways: i) by combining the two observations (4.3) and (4.4) in one and then using equation (4.1), or ii) by the two-step process consisting of first obtaining the Bayesian estimate of **x** using one of the observations and the Fisher estimate of **x** using the other observation, and then considering these two estimations as new observations and obtaining the corresponding Fisher estimate.

Notice that to be able to use the two-step approach in the Bayesian case, the state **x** is considered at the same time as a random variable (with a priori information available) and as an unknown constant. This is interpreted as that the a priori information is used only once while the Fisher model concepts are used for the rest of the estimation. Figure 4.2 depicts schematically the two cases.

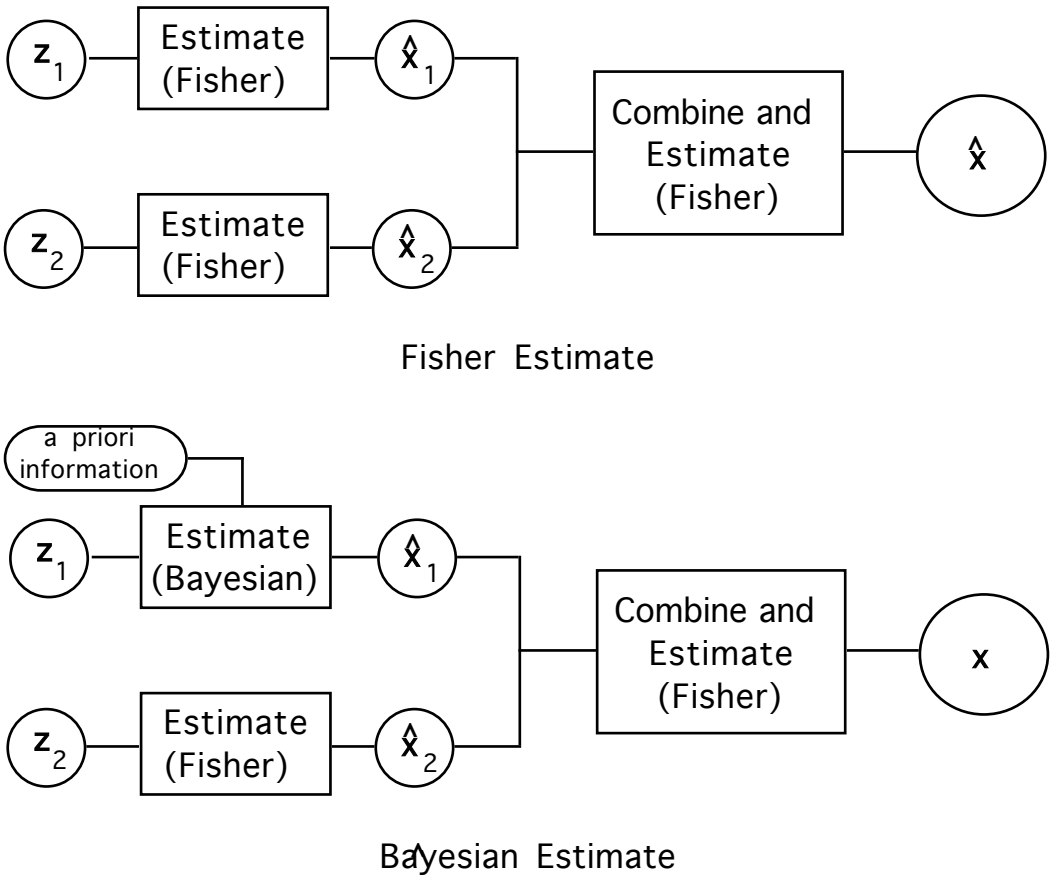


Figure 4.2. Estimation from two observations for the Fisher and Bayesian models.

The Kalman filter will be derived as a Bayesian estimate of the state **x** when two observations are available. These two observations will be the measurements themselves at time k+1 and the estimation of the state at time k+1 obtained by propagation of the last estimate at time k using a linear model that describes the dynamics of the system. It is this last observation, based on an estimate of the system at an earlier time, that will carry all the a priori information.

4.2. The Kalman filter

Let $\mathbf{x}(k)$ be a dynamic system that can be represented by the state-transition model

$$\mathbf{x}(k+1) = \Phi(k)\mathbf{x}(k) + \Gamma(k)\mathbf{w}(k) \tag{4.7}$$

where $\Phi(k)$ is the transition matrix between time steps k and k+1 and $\Gamma(k)\mathbf{w}(k)$ represents the model error with zero mean and covariance matrix $\Gamma(k)\mathbf{Q}(k)\Gamma^T(k)$. For the sake of simplicity, no deterministic forcing term is considered, although it would be straightforward to include it in the equations. In any case, it does not have any influence on the estimation error covariance matrix.

Observations will be made according to the observation equation

$$\mathbf{z}(k) = \mathbf{H}(k)\mathbf{x}(k) + \mathbf{v}(k)$$

The Kalman filter is made up of two basic procedures, one-step prediction of $\mathbf{x}(k+1)$ given the observations $\mathbf{z}(1)$, $\mathbf{z}(2)$, …, $\mathbf{z}(k)$ and filtering of this prediction by means of the observation $\mathbf{z}(k+1)$.

If the conditional mean of $\mathbf{x}(k)$ given observations $\mathbf{z}(1)$, $\mathbf{z}(2)$, …, $\mathbf{z}(k)$ is known and represented by $\hat{\mathbf{x}}(k|k)$, with error covariance

$$\Sigma(k|k) = E\{(\mathbf{x}(k) - \hat{\mathbf{x}}(k|k))(\mathbf{x}(k) - \hat{\mathbf{x}}(k|k))^T\}$$

the minimum variance one-step prediction, $\hat{\mathbf{x}}(k+1|k)$ can be obtained by taking conditional means in equation (4.7)

$$\hat{\mathbf{x}}(k+1|k) = \Phi(k)\hat{\mathbf{x}}(k|k)$$

The term in $\mathbf{w}(k)$ vanishes as a consequence of the whiteness and zero mean property of $\mathbf{w}(k)$.

The mean square forecasting error is given by

$$\begin{aligned}
\Sigma(k+1|k) &= E\{(\mathbf{x}(k+1) - \hat{\mathbf{x}}(k+1|k))(\mathbf{x}(k+1) - \hat{\mathbf{x}}(k+1|k))^T\} \\
&= E\{[\Phi(k)(\mathbf{x}(k) - \hat{\mathbf{x}}(k|k)) + \Gamma(k)\mathbf{w}(k)] \times \\
&\qquad [\Phi(k)(\mathbf{x}(k) - \hat{\mathbf{x}}(k|k)) + \Gamma(k)\mathbf{w}(k)]^T\} \\
&= \Phi(k)E\{(\mathbf{x}(k) - \hat{\mathbf{x}}(k|k))(\mathbf{x}(k) - \hat{\mathbf{x}}(k|k))^T\}\Phi^T(k) \\
&\qquad + \Gamma(k)E\{\mathbf{w}(k)\mathbf{w}^T(k)\}\Gamma^T(k) \\
&= \Phi(k)\Sigma(k|k)\Phi^T(k) + \Gamma(k)\mathbf{Q}(k)\Gamma^T(k)
\end{aligned} \tag{4.8}$$

At this point (and after the observation $\mathbf{z}(k+1)$ has been made), we have an estimate $\hat{\mathbf{x}}(k+1|k)$ of the state vector at step $k+1$ that has been made using all the a priori information available (it is, in fact, the conditional mean of the state at $k+1$ given all the observations at previous instants) and an observation at time $k+1$. Recalling the results from the two-step estimation, the Bayesian estimation from two observations can be regarded as a Fisher estimation of two "observations" (estimations), one of them made using the a priori information, and the other one without assuming any information. This is the case here, the estimate $\hat{\mathbf{x}}(k+1|k)$ carries all the a priori information and the observation $\mathbf{z}(k+1)$ is considered as the observation of an unknown constant. Therefore, the minimum variance linear unbiased estimate, $\hat{\mathbf{x}}(k+1|k+1)$, of $\mathbf{x}(k+1)$ can be obtained as the Fisher estimate of the two "observations":

$$\hat{\mathbf{x}}\ (k+1|k) = \mathbf{x}(k+1) - \delta\,\mathbf{x}(k+1|k)$$
$$\mathbf{z}(k+1) = \mathbf{H}(k+1)\mathbf{x}(k+1) + \mathbf{v}(k+1) \qquad (4.9)$$

Augmenting the observation vector,

$$\tilde{\mathbf{z}}(k+1) = \mathbf{H}(k+1)\tilde{\mathbf{x}}(k+1) + \tilde{\mathbf{v}}(k+1)$$

with

$$\tilde{\mathbf{z}}(k+1) = \begin{bmatrix} \hat{\mathbf{x}}\ (k+1|k) \\ \cdots\cdots\cdots \\ \mathbf{z}(k+1) \end{bmatrix}$$
$$\tilde{\mathbf{H}}(k+1) = \begin{bmatrix} \mathbf{I} \\ \cdots\cdots \\ \mathbf{H}(k+1) \end{bmatrix}$$
$$\tilde{\mathbf{v}}(k+1) = \begin{bmatrix} -\delta\,\mathbf{x}(k+1|k) \\ \cdots\cdots\cdots\cdots \\ \mathbf{v}(k+1) \end{bmatrix} \qquad (4.10)$$

$$\tilde{\mathbf{R}}(k+1) = E\{\tilde{\mathbf{v}}(k+1)\tilde{\mathbf{v}}^{T}(k+1)\}$$
$$= \begin{bmatrix} \Sigma\ (k+1|k) & \vdots & \mathbf{0} \\ \cdots\cdots\cdots & \vdots & \cdots\cdots \\ \mathbf{0} & \vdots & \mathbf{R}(k+1) \end{bmatrix}$$

a Fisher estimation yields the estimate

$$\tilde{\mathbf{x}}(k+1|k+1) = \Sigma\,(k+1|k+1)\tilde{\mathbf{H}}^{T}(k+1)\tilde{\mathbf{R}}^{-1}(k+1)\tilde{\mathbf{z}}(k+1)$$

$$\Sigma\,(k+1|k+1) = [\tilde{\mathbf{H}}^{T}(k+1)\tilde{\mathbf{R}}^{-1}(k+1)\tilde{\mathbf{H}}(k+1)]^{-1}$$

Substitution of equations (4.10) transforms the estimate into the dynamic Kalman filter

$$
\begin{aligned}
\hat{\mathbf{x}}(k+1|k+1) &= \boldsymbol{\Sigma}(k+1|k+1)[\mathbf{H}^T(k+1)\mathbf{R}^{-1}(k+1)\mathbf{z}(k+1) \\
&\quad + \boldsymbol{\Sigma}^{-1}(k+1|k)\boldsymbol{\Phi}(k)\hat{\mathbf{x}}(k|k)] \\
\hat{\mathbf{x}}(k+1|k) &= \boldsymbol{\Phi}(k)\hat{\mathbf{x}}(k|k) \\
\boldsymbol{\Sigma}(k+1|k) &= \boldsymbol{\Phi}(k)\boldsymbol{\Sigma}(k|k)\boldsymbol{\Phi}^T(k) + \boldsymbol{\Gamma}(k)\mathbf{Q}(k)\boldsymbol{\Gamma}^T(k) \\
\boldsymbol{\Sigma}(k+1|k+1) &= [\mathbf{H}^T(k+1)\mathbf{R}^{-1}(k+1)\mathbf{H}(k+1) \\
&\quad + \boldsymbol{\Sigma}^{-1}(k+1|k)]^{-1}
\end{aligned}
\qquad (4.11)
$$

As it is formulated, the Kalman filter is a set of recursive equations that cannot be solved without a set of initial conditions. The initial conditions needed are i) the best estimate of the state variable at time zero and before any measurement has been performed, which is the expected value, and ii) the mean square error of this estimate, that, when the estimate is the (known) expected value, turns out to be the covariance function of the process at time zero.

$$\hat{\mathbf{x}}(0|0) = \mathbf{m}$$

$$\boldsymbol{\Sigma}(0|0) = \boldsymbol{\Psi}$$

Before proceeding, it is important to notice that the estimation error covariance matrix obtained in equation (4.11) is independent of the observations. This means that the observations not need to be made in order to know the covariance of the estimation error of the system. The estimation error covariance matrix depends only on the geometric location of the observations but not on the actual values of these observations. This fact is particularly interesting in network design when some measure of the estimation error covariance matrix is used to assess the performance of the network, for the network not need to be operative in order to infer its performance. The network can be designed without the need of taking one sample! On the other hand this means that the estimations obtained using the Kalman filter do not take into account the reliability of the observations. It would be interesting if the estimation error covariance were dependent on the observations, such that if one of the measurements has a large measurement error or comes from one tail of the probability distribution of **x** then the

estimation error should be larger than if the measurement comes from the center of the distribution or the measurement error is very small.

After some algebraic manipulations the Kalman filter can be rewritten in several other forms that may be more convenient in certain cases. For instance, in the case of perfect measurements, the matrix **R** is a zero-matrix and the estimation covariance cannot be calculated using the previous formulations. In such a case, the following equation is more convenient

$$\begin{aligned}\boldsymbol{\Sigma}(k+1|k+1) = \boldsymbol{\Sigma}(k+1|k) - \boldsymbol{\Sigma}(k+1|k)\mathbf{H}^T(k+1) \\ \times [\mathbf{R}(k+1) + \mathbf{H}(k+1)\boldsymbol{\Sigma}(k+1|k)\mathbf{H}^T(k+1)]^{-1} \\ \times \mathbf{H}(k+1)\boldsymbol{\Sigma}(k+1|k)\end{aligned} \tag{4.12}$$

Besides the fact that equation (4.12) allows perfect measurements, it may be computationally more efficient, since it involves the inversion of an $m \times m$ matrix whereas equation (4.11) requires the inversion of an $n \times n$ matrix, which in general is much larger; on the other hand equation (4.12) requires a larger number of $n \times n$ matrix multiplications. Equation (4.12) shows that the estimation error covariance is equal to the forecasting error covariance less a positive definite matrix which is function of the measurement errors and the sample locations. Consequently, observations help to reduce forecasting error.

After some more matrix manipulation the filter may be rewritten in its most popular form:

$$
\begin{aligned}
\hat{\mathbf{x}}(k+1|k+1) &= \hat{\mathbf{x}}(k+1|k) + \mathbf{K}(k+1)[\mathbf{z}(k+1) \\
&\quad - \mathbf{H}(k+1)\hat{\mathbf{x}}(k+1|k)] \\
\hat{\mathbf{x}}(k+1|k) &= \mathbf{\Phi}(k)\hat{\mathbf{x}}(k|k) + \mathbf{\Lambda}(k)\mathbf{U}(k) \\
\mathbf{K}(k+1) &= \mathbf{\Sigma}(k+1|k)\mathbf{H}^T(k+1)[\mathbf{R}(k+1) \\
&\quad + \mathbf{H}(k+1)\mathbf{\Sigma}(k+1|k)\mathbf{H}^T(k+1)]^{-1} \\
&= \mathbf{\Sigma}(k+1|k+1)\mathbf{H}^T(k+1)\mathbf{R}^{-1}(k+1)
\end{aligned}
\tag{4.13}
$$

$$
\mathbf{\Sigma}(k+1|k+1) = [\mathbf{I} - \mathbf{K}(k+1)\mathbf{H}(k+1)]\mathbf{\Sigma}(k+1|k)
$$

This is a very intuitive form where the filtered estimate is expressed as the forecasted estimate plus a linear combination of the departures of the observations from their predicted values. The coefficients of this linear combination are the Kalman gain matrix $\mathbf{K}(k)$. Notice that the possibility of a nonzero deterministic forcing term $\mathbf{\Lambda}(k)\mathbf{U}(k)$ has been included in equation (4.13). The only effect of this term is in the forecasting estimate. The difference $\mathbf{z}(k+1)$-$\mathbf{H}(k+1)\hat{\mathbf{x}}(k+1|k)$ is called the innovation sequence in the sense that it expresses the discrepancy (innovation) between prediction and observation.

As a final comment, one must realize that no assumption on the probability distribution of the state variable has been made. In particular, normality is not required. Only the mean estimate and the initial estimation covariance are needed. If **x** and **v** are jointly normal then the linear estimation provided by the Kalman filter is optimal.

5. AN EXAMPLE OF DYNAMIC NETWORK DESIGN

Dimensionality problems and computational cost restricted me to a focus on one-dimensional problems. However, the methodology presented here is applicable to any dimension of problem, with the only difference coming from the building of matrix $\mathbf{\Phi}'$ in equations (3.8) and (3.14). This is, as a matter of fact, the only difference in the application of this methodology to any other linear model for which a state-transition model of the form of (3.5) can be written.

There are other reasons for me to choose one-dimensional theoretical case studies. They can provide some understanding of how Kalman filters work in the process of forecasting and filtering. At the same time, many different simple cases can be solved to study the role of changes in aquifer parameters. The results for such simple cases might be extrapolated to higher dimensional problems and might help to find a more efficient way than the one used here to deal with real areal two-dimensional cases.

5.1. Statement of the network design problem

The objective of all network design problems is to somehow minimize the uncertainty that the hydrologist has about *something* in the system. There are many ways of interpreting this sentence and some of them were mentioned in the literature review:

- uncertainty about piezometric heads, solute concentration or any other state variable at each and every point of the system,
- same as above but only at a certain region of the system,
- uncertainty about the amount of contaminant within the system or within a portion of it,
- uncertainty about the shape of a contaminant plume,
- uncertainty about peak values,
- uncertainty about the parameters governing the system,
- uncertainty about the detection of violations of standards,
- uncertainty about the necessary cost to clean up an aquifer,

and as many others as one can imagine based on the facts that real systems vary continuously in space in a way that is impossible to reproduce, that only a few values of the state variables

and/or the parameters are available, and that even those few values are imprecise due to measurement errors.

In this study, the network design problem will focus on what I will call the global error or global uncertainty over space and time of the system. Based on the fact that the Kalman filter provides a way to obtain an estimation error covariance matrix at each point in space and at each moment in time, before and after any measurement is made, independently of the actual measurements, all the sampling schemes will be compared over the same time period and the mean value of the estimation error variance over space and time will be computed. This mean global estimation error will be used as an index to evaluate the performance of the scheme. This index is more appropriate when the objective of the study depends on a global evaluation of the phenomenon under study, for instance, to monitor contaminats from a regional source (as it is the case of nitrite contamination from fertilizers used in agriculture).

The only constraints in the design are a finite set of spatial locations where samples may be gathered and the total number of samples that will be gathered at all those locations. Both must be known a priori.

It is assumed that an estimation of the state of the system as well as its estimation error is known at time zero. The parameters are perfectly known and the model of the system is also perfect. Measurement errors are allowed.

To illustrate the meaning of the global error associated with a given sampling scheme, an example is given. The flow model, which is described in detail in the following section, is used. In this case the network is used to obtain measurements of piezometric head in an unidimensional aquifer with constant heads at both ends. This example refers to the evaluation of the global uncertainty in head associated with a design consisting of taking a single measurement at time t=0.5T and location x =0.5L .

At time zero the error covariance matrix is known, and in particular the error variance at each location. In Figure 5.1, the spatial distribution of error variances at time zero is shown.

The same uncertainty is assigned to every point and it is assumed that there is no spatial correlation between initial estimation errors. The objective is to evaluate the mean estimation error variance over a given period of time T , and over the length of the aquifer L , provided that a measurement is made at t=0.5T and at location x=0.5L. For this purpose the error variance must be known at each time step and at each location from t=0 until t =1.0T. This is done by propagating the error covariance at t=0 until before a measurement is made (t=0.5T^-) using the third equation in (4.11).

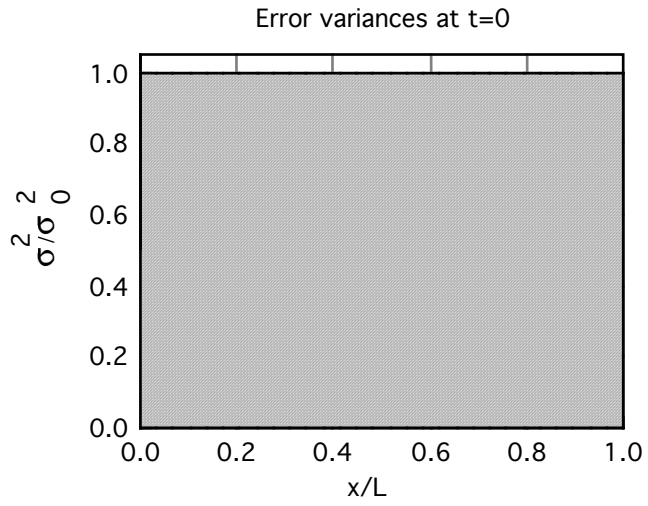


Figure 5.1. Variance distribution at time zero (both axis are normalized, L is the total length and σ_0 is the initial error variance)

At times t=0.3 T and t=0.5 T^- the variance distribution has evolved as shown in Figure 5.2. At this time a measurement is taken and the forecasted (prior) variance distribution is updated using the fourth equation in (4.11) to come out with the filtered (posterior) variance distribution shown in Figure 5.3. This filtered variance is finally propagated in time until t=1.0T , since no other measurement is to be made.

For each time step the mean variance can be calculated by integration of curves similar to those shown in Figures 5.1—5.3 and a final curve is obtained that reflects the evolution in time of the mean estimation error variance (Figure 5.4). The integral over time of the curve in Figure 5.4 will provide a measurement of the mean estimation error variance with which the

system has been known over the entire period *T*. This value will be used as an indicator of the performance of the sampling scheme.

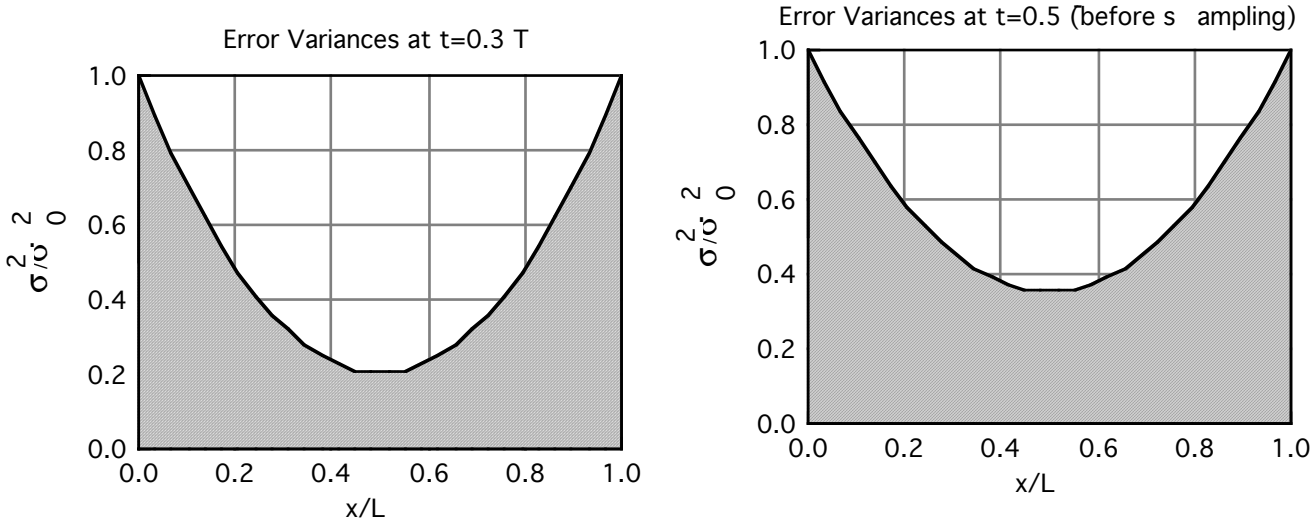


Figure 5.2. Forecasted error variances at times 0.3 *T* and 0.5 *T*

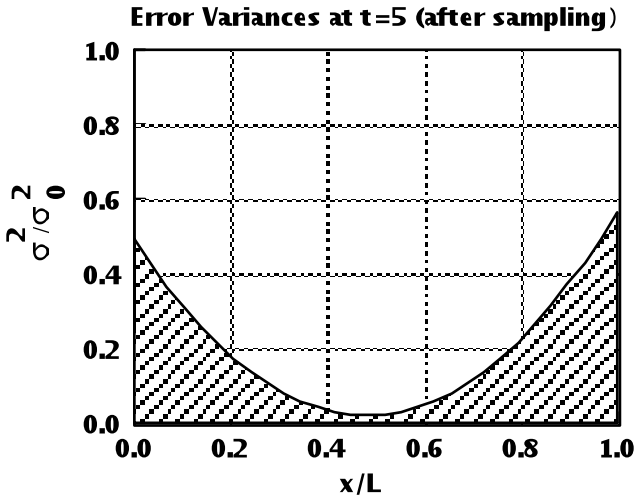


Figure 5.3. Change in variance after a measurement is done at *x/L*=0.5

In particular, for the example shown in Figure 5.4, the integral of this curve evaluates to 0.30, which may be interpreted as meaning that the state of the system has been known at each point and at each time step, with an estimation error, on average, of 30% of the initial estimation error.

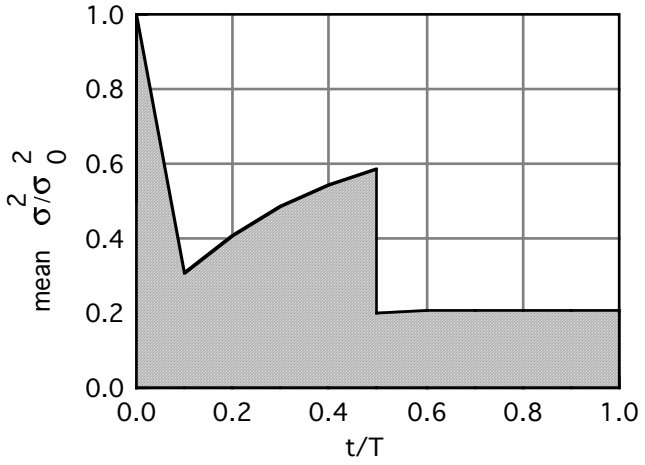


Figure 5.4. Evolution in time of the spatially averaged error variance when a measurement is made at t= 0.5 T at location x= 0.5L . (Refer to next section for a better understanding of the meaning of this figure).

The analytical formalism that gives the global error over a period of time and over a certain region of the aquifer is

$$\sigma^2 = \frac{1}{T \cdot A}\left(\int_{t_0}^{t_1^-}\int_{\Omega} \Sigma\,(t;\mathbf{x})\,d\Omega\,dt + \int_{t_1^+}^{t_2^-}\int_{\Omega} \Sigma\,(t;\mathbf{x})\,d\Omega\,dt + \cdots \right.$$
$$\left. \int_{t_{n-1}^+}^{t_n}\int_{\Omega} \Sigma\,(t;\mathbf{x})\,d\Omega\,dt \right) \qquad (5.1)$$

where $\boldsymbol{\Sigma}(t; \mathbf{x})$ is the estimation variance at time t and location $\mathbf{x}$, A is the area (or length) of the domain of interest (it may be the entire aquifer or just part of it), T is the interval of time in which the sampling scheme is evaluated, Ω is the domain, t_0 is the initial time and t_n=T ; t_1, …, t_{n-1} are sampling times on which the error estimation variance is not a continuous function. The superscripts - and + indicate before and after a sampling has been made, respectively.

5.2. One-dimensional groundwater flow model

The Kalman filter technique has been evaluated in a one-dimensional aquifer, with constant transmissivity, constant heads at both ends, and no sinks or sources. The initial value of the

estimation error covariance matrix $\mathbf{\Sigma}(0|0)$ was assumed to be of the form $\sigma^2_0\mathbf{I}$ where $\mathbf{I}$ is the identity matrix and σ^2_0 is a large number. Therefore, no spatial correlation between the initial errors was assumed. (This lack of correlation in the initial estimation error covariance matrix produces a distinctive behavior at the early stage of the forecasting process, as will be seen later). The best estimate of the initial piezometric heads $\mathbf{x}(0|0)$ was assumed to be zero everywhere except at the constant head boundary at x=0 where the piezometric head is estimated to be one.

Since neither the forecasted estimation covariance matrix nor the filtered estimation covariance matrix are dependent on the observed head values or on the head values predicted by the model, and since the measure of uncertainty associated with each sampling scheme depends only on the estimation covariance matrices, the initial heads and the evolution of heads in time do not have any influence on the decision-making process. They are given as a point of reference but it should be noted that any other set of initial head values would have given these same results, provided that the initial estimation covariance matrix was the same.

Although a numerical model was built and tested to deal with two-dimensional problems, the simplifications introduced in this one-dimensional case permit building matrix $\mathbf{\Phi}'$ in equation (3.8) in a very straightforward way as a function of only one coefficient.

$$\mathbf{\Phi}' = \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots & 0 \\ \xi_1 & -2\xi_1 & \xi_1 & 0 & \cdots & 0 \\ 0 & \xi_1 & -2\xi_1 & \xi_1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 \end{bmatrix}; \qquad \xi_1 = \frac{T}{S(\Delta x)^2} \tag{5.2}$$

where T is the transmissivity, S is the specific storage and Δx is the discretization size. Notice that the first and last rows are zeroes since they correspond to a constant head boundaries for which dh/dt is zero. An implicit scheme is used to approximate the time derivative, so that equation (3.11) is used to obtain the final transition matrix

$$\Phi = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ -\xi & 1+2\xi & -\xi & 0 & \cdots & 0 \\ 0 & -\xi & 1+2\xi & -\xi & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix}; \quad \xi = \frac{T}{S(\Delta x)^2}\Delta t \qquad (5.3)$$

The aquifer was discretized into 30 cells and the time interval was chosen such that the final head distribution was 99% of the steady state distribution. The value for the coefficient ξ_1 was 36, although, as is explained later, the results are applicable for any other value.

Before proceeding to the network design itself, it is worthwhile to take a look at how the model will propagate the initial estimation covariance through time without considering any measurement. Again, recall that the actual values of the piezometric heads do not have any influence at all on the estimation covariance results. They do have influence on the estimation of the piezometric heads themselves, though. For this reason the choice of the length of the simulation period is somewhat arbitrary since it is based on a particular set of initial conditions. The same results, regarding the estimation covariance matrices, would have been obtained had steady state already been achieved at time zero.

Figure (5.5) shows the state of the system at different moments in time just as a reference as to how the system evolves as times goes by.

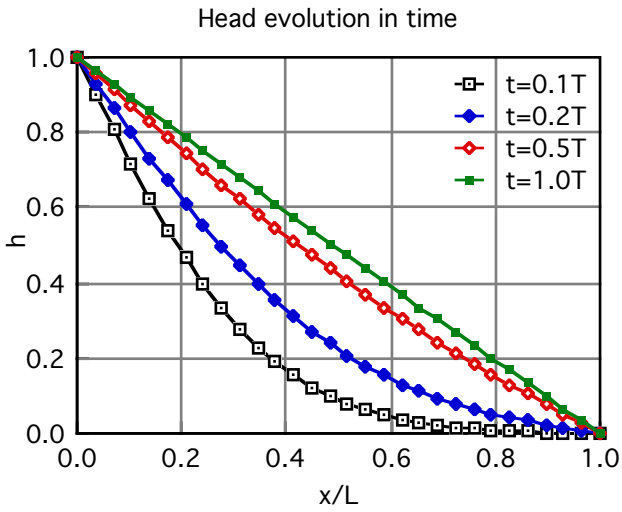


Figure 5.5. Evolution of heads in time when the initial heads are zero except at x=0 where the head is 1. (Both ends are constant head boundaries).

The next figures show several possible views of the estimation covariance matrix as it changes in time. Remember that the estimation covariance at time zero is the identity matrix multiplied by a coefficient and that there is no spatial correlation between errors at different locations. Figure 5.6 shows how correlation is built into the estimation covariance matrix as time passes.

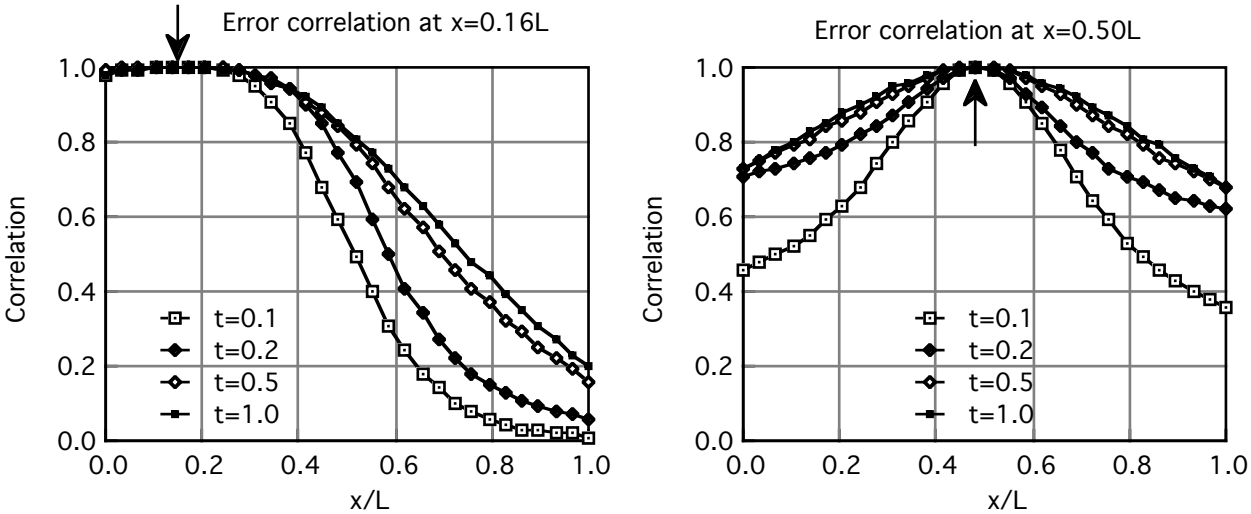


Figure 5.6. Evolution in time of the spatial correlation between estimation errors at x=0.16L and x=0.5L and the rest of the aquifer.

For x=0.16L the estimation error soon becomes totally dependent on the errors in an area around it. The correlation with errors in the heads at the other end of the system increases, although it seems that a certain degree of independence will remain even when steady state has

been reached. For x=0.5L something similar happens, but the degree of correlation achieved with the surrounding points is not as large as it is at x=0.16L . The results for x=0.83L , although not shown, are the mirror image of the x=0.16L case. This can be explained from the fact that in steady state the heads are dependent only on the constant head boundary values and therefore so are the estimation errors. Thus, since the estimation errors at the boundaries determine the estimation error everywhere, the closer to a boundary, the higher the effect of that boundary on the estimation error, and consequently the higher the correlation. At the middle point, the estimation error depends equally on the estimation error at both ends but not totally on either of them. It is also important to note the big change in the correlation structure that is introduced just after the first time step and the smoother variation later in time.

In Figures 5.7 the change in correlation after a measurement has been made at the end of the simulation period is depicted.

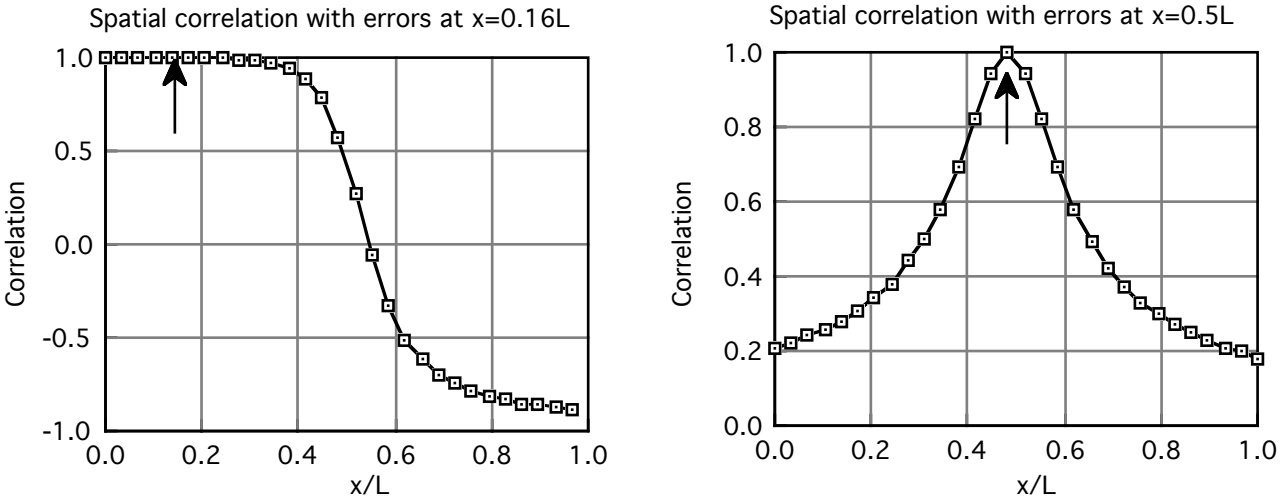


Figure 5.7. Spatial correlation between the estimation errors at x=0.16L and x=0.5L and the estimation errors at the rest of the aquifer after a measurement has been done at x=0.5L and steady state has been achieved.

The results of these Figures have again a clear physical interpretation. After steady state has been achieved and an almost exact measurement is made at x=0.5L (a measurement error of only 2% of the initial estimation error was considered) the correlation between the estimation error at x=0.16L and the errors at locations between 0.5L and 1.0L must be negative. The reason is that the system state must be a constant head gradient between both ends. Therefore, if an exact measurement has been made at x=0.5L and the estimation at x=0.16L is above the

real value, the estimate of the gradient would be higher than the real one and the estimation error at x=0.83L must be of the same order of magnitude but with opposite sign (and almost completely correlated). The correlation with the estimation error at x=0.5L would have been zero if no measurement error had been introduced. In any case, the variance itself at that point is very small, as can be seen in Figure 5.8, so the correlation does not have much meaning in this case.

Whereas the evolution of the spatial correlation seems to have a clear physical interpretation, it is not so in the case of the evolution of the absolute value of the variances themselves (Figure 5.8).

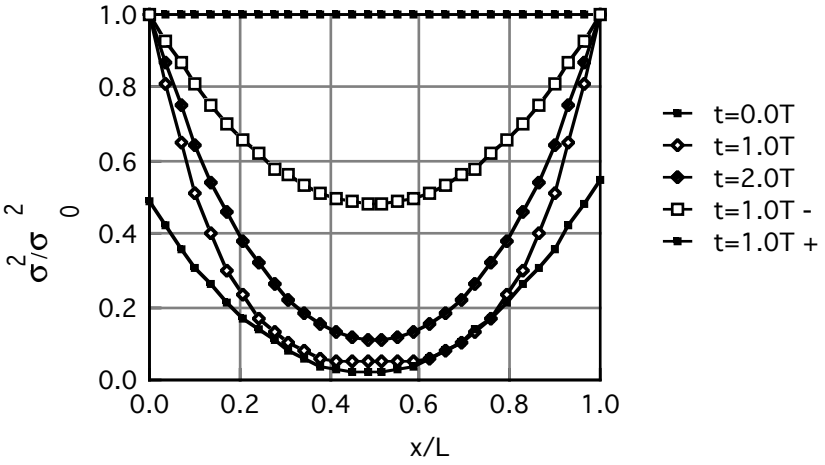


Figure 5.8. Evolution of estimation variance in time relative to the estimation variance at time zero. (The last two curves in the legend refer to the variance before and after an almost exact measurement has been done at x=0.5L.)

The estimation variance at steady state will depend only on the estimation variance at both ends and it is easy to find its analytical expression. The head distribution at steady state is given by

$$h(x) = h_0\left(1 - \frac{x}{L}\right) + h_L \frac{x}{L}$$

where h_0 and h_L are the heads at both ends and L is the length of the aquifer. The estimate and the estimation variance, assuming that there is no correlation between h_0 and h_L will be given by

$$\hat{h}(x) = \hat{h}_0(1 - \frac{x}{L}) + \hat{h}_L \frac{x}{L}$$

$$\sigma^2_h(x) = \sigma^2_{h_0}(1 - \frac{x}{L})^2 + \sigma^2_{h_L}(\frac{x}{L})^2$$

The parabolic shape of the final estimation variance is indeed obtained by the forecasting model (Figure 5.8). But the intermediate states are more difficult to interpret. Notice that the estimation variance after the first time step reduces dramatically from the horizontal line that represents the initial state. This is to say that no matter how bad the estimate of the initial state the predicted results will be very accurate in the early stages of the simulation. This behavior stems from the form of the initial correlation matrix and the fact that no spatial correlation was considered. Indeed, if an exponential correlation is considered in the initial estimation errors, with integral scale equal to one third of the length of the aquifer, the variance does not decrease so much at the early stages. (Figure 5.9) See also Figures 5.12 and 5.13 discussed below. Notice that the correlation introduced is such that the errors in the estimation of the constant head boundaries remain uncorrelated.

In both Figures 5.8 and 5.9 the results for steady state are the same but the transient results are completely different. This points out the large role that the initial estimation error covariance matrix has on the results for the transient situation. For the extreme case, with an infinite integral scale in the initial estimation covariance matrix , that is, perfect correlation between any two points in the aquifer, the initial estimation covariance matrix remains constant in time since all the errors depend on the error committed in one of the head boundaries and the model cannot improve its estimation unless a measurement is done. However, I think that the initial estimation error covariance that best represents a lack of knowledge at time zero is the diagonal matrix with a large determinant, although it has been seen that this covariance matrix produces some non-intuitive results at the beginning of the simulation until some spatial correlation is built into the estimation errors by the model. On the other hand, we may have a total lack of knoweledge about the state of the system but the fact that we know the physics

underlying the phenomenon should be enough information to build a different initial covariance matrix.

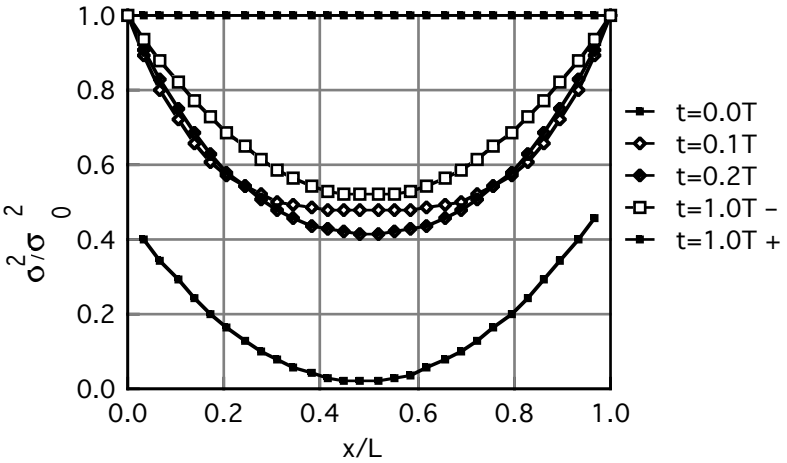


Figure 5.9. Evolution of estimation variance in time relative to the estimation variance at time zero. The initial error estimation variance had an exponential spatial correlation with integral scale equal to one third of the length of the aquifer. (The last two curves in the legend refer to the variance before and after an almost exact measurement has been done at x=0.5L.)

The reduction provided by the measurement at $x=0.5L$ can be explained in the following way: the steady state depends only on the values of the constant head boundaries, what can be interpreted as if the system had two degrees of freedom. If an almost exact measurement is made, the degrees of freedom reduce to one and the estimation error at the boundaries reduces by one half.

Next, a simple case of network design to determine which are the best time and location to take one measurement is addressed. The measure of uncertainty given by (5.1) is used to discriminate between the possible choices. The modeling problem is fully determined by two parameters ξ_1 in (5.2) and the length of the simulation period. The first one was chosen to be ξ_1=36 (units of time)$^{-1}$ and the second one was chosen such that the piezometric heads at the end of the simulation were 99% of the steady state heads. This value happened to be equal to 1 unit of time. The number of time steps for which the simulation was carried out is 100. The results obtained are valid for any other situation where the product ξ_1×(simulation time) is the same and the constraints in number of samples and frequency of sampling are the same, as can be concluded from the form of the matrix $\mathbf{\Phi}$ in (5.3).

The objective was to find the sample that provides the least global uncertainty to be taken among locations 0.066L, 0.16L, 0.33L, 0.5L, 0.66L, 0.83L or 0.93L and times 0.0T, 0.1T, 0.2T, 0.3T, 0.4T, 0.5T, 0.6T, 0.7T, 0.8T, 0.9T or 1.0T, (T being the total simulation time). This gives a total of 7 locations × 11 instants, that is, 77 possible sampling schemes. Remember that for each of these sampling schemes a complete simulation must be performed in order to compute the uncertainty associated with it. This is only to locate the best "one sample", if the best "two samples" were to be found the number of possible schemes to be analyzed rises to 2926. The procedure used in this work to find the best scheme is an exhaustive search. This can be carried out for small systems and for not too many possible sampling choices. A way to eliminate possible candidates without having to perform the simulation needs to be found in order to make this methodology feasible in practice.

The results for this network design problem are spatially symmetric, as might be expected from the conjunction of a number of factors: the matrix $\mathbf{\Phi}$ in (5.3) is symmetric, the initial estimation variances are constant and the forecasted and filtered covariance matrices do not depend on the heads. This means that the same uncertainty is associated with measurements at time t at any two points equally distant from the ends.

The uncertainty corresponding to each sample location and each sample instant computed using equation (5.1) is shown in Figures 5.10 and 5.11.

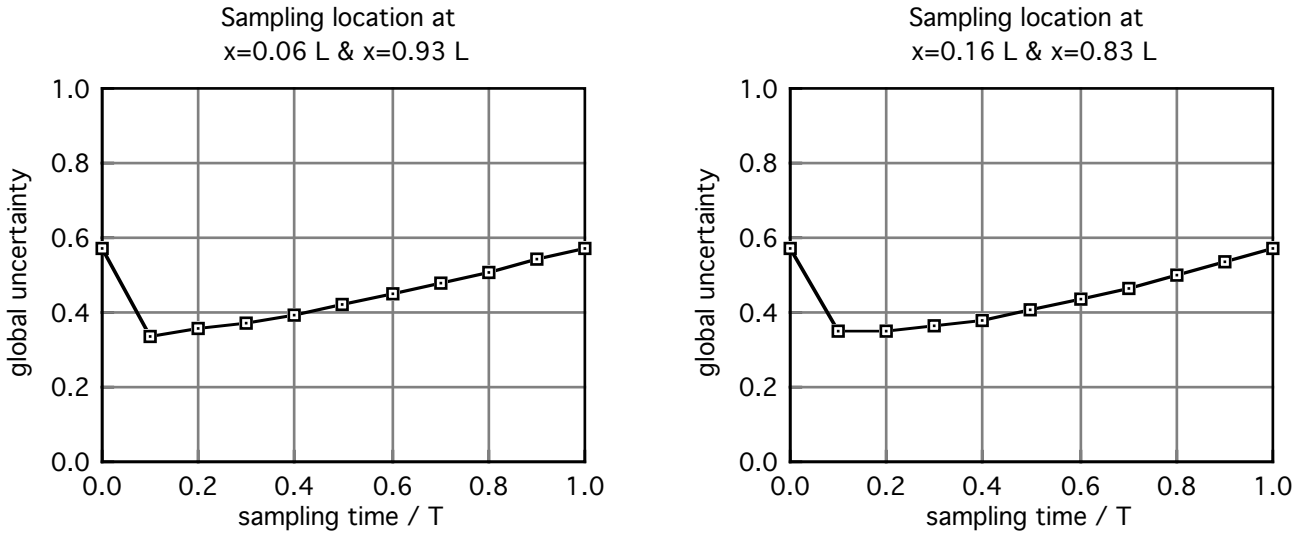


Figure 5.10 (Continued on next page)

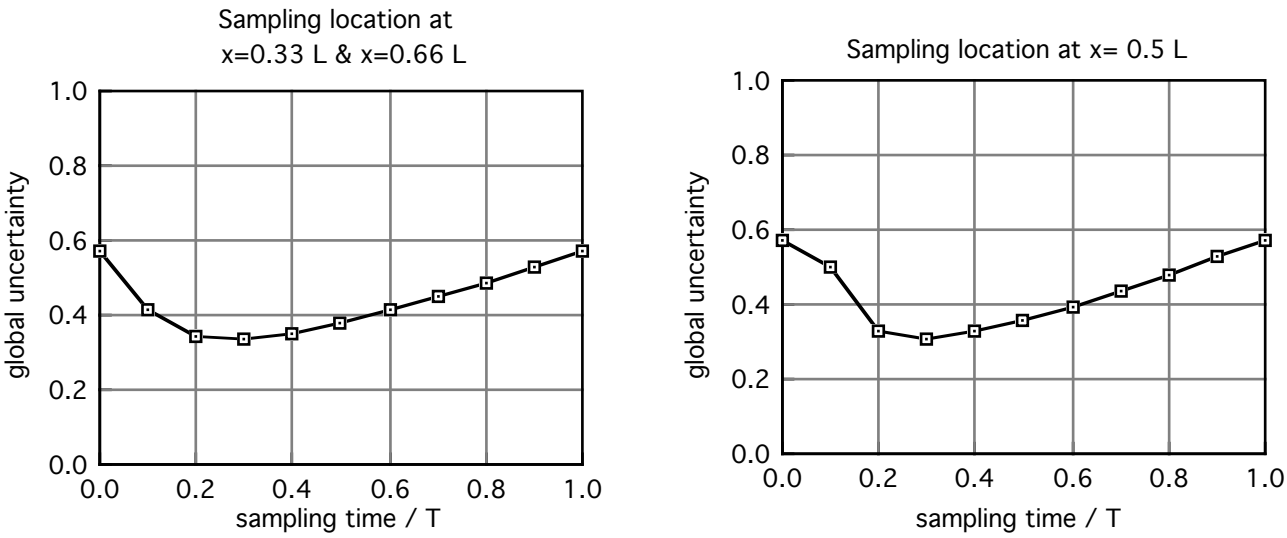


Figure 5.10. Each dot in the plots represents the mean uncertainty in the knowledge of the system during the total duration of the simulation when a sample is taken at the time and location indicated.

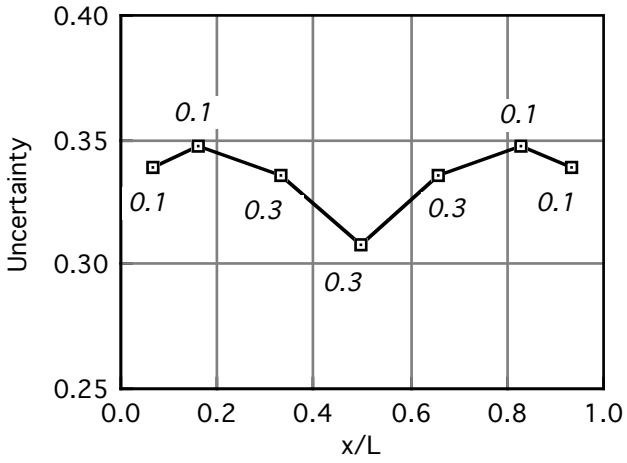


Figure 5.11. For each location the best sampling time is chosen from Figure 5.10. (The time relative to the total simulation time is shown in italics)

From these results it is concluded that the best location to take a sample is at x=0.5L and at t=0.3T . Sampling at this time and place results in a mean estimation error of 31% of the initial estimation error whereas sampling at x=0.16L will give, at the best, a mean estimation error of almost 34% of the initial estimation error. This 10% difference between the sampling schemes may not be enough to justify the computational effort, since a good practitioner would probably have chosen to sample at the centerpoint and at some early time, given the expected transient behavior.

Given that for a particular groundwater flow model the results depend only on the initial estimation covariance matrix, and given the unusual behavior observed in the evolution of the variance over time, the same network design problem was solved for an initial estimation covariance matrix with a spatial correlation structure of the exponential type with integral scale equal to one third of the length of the aquifer. The results for this new problem are shown in Figures 5.12 and 5.13

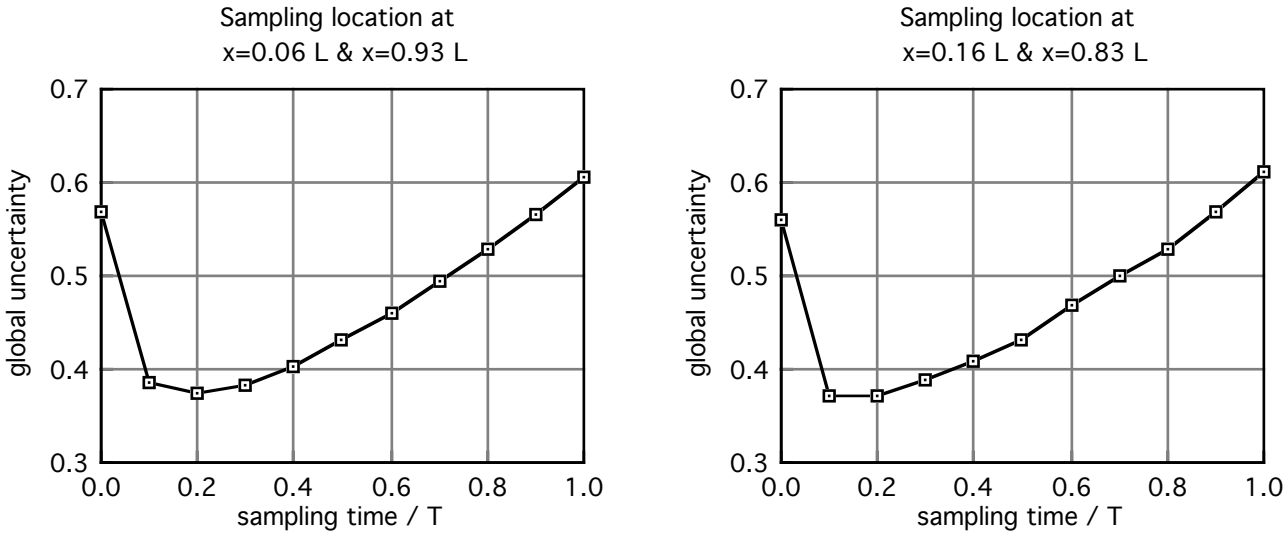


Figure 5.12. (Continued on next page)

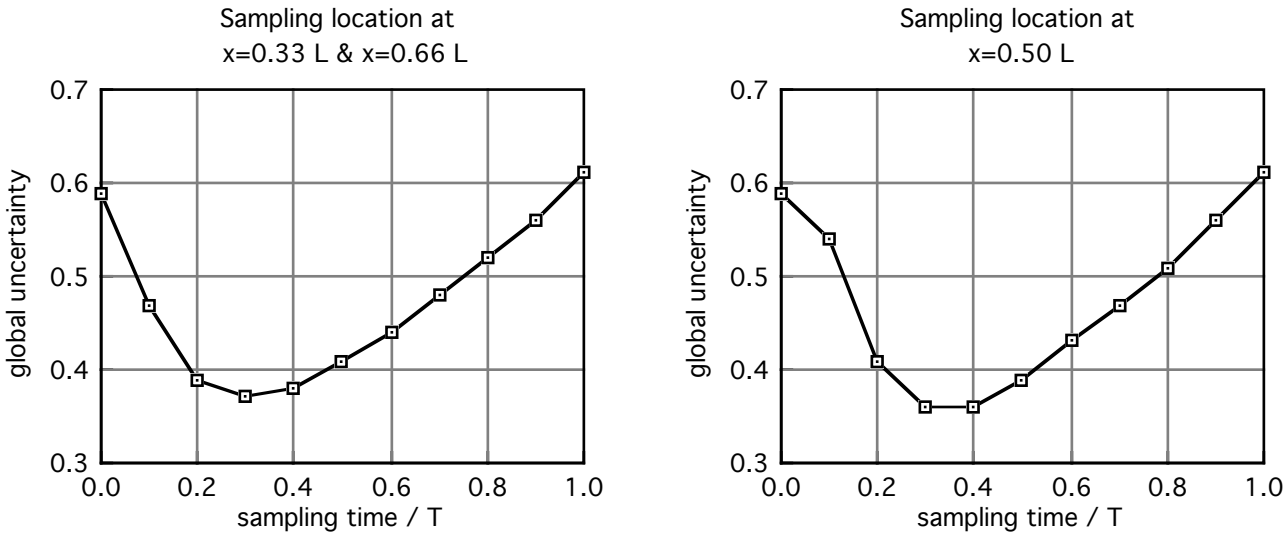


Figure 5.12. Global uncertainty at each point and each location when the initial estimation covariance matrix includes spatial correlation.

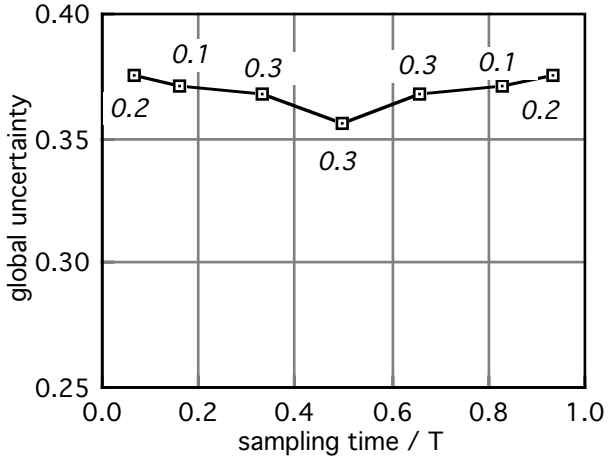


Figure 5.13. For each location in Figure 5.12 the best sample has been chosen (shown in italics).

There are some differences between the results for the two different initial conditions, but no major ones. The best sampling again happens to be at x=0.5L and at t=0.3T although in the latter case the mean estimation error is larger, 35.6% of the initial error, and the differences in global uncertainty between the best samples at different locations are in a smaller range.

The location of the best sampling was expected and could have been predicted in advance, but not so the time for the best sampling at any of the locations. I cannot find a clear physical explanation for why near the boundaries it is best to take the measurement early in time, and why the best time increases with the distance to the boundaries. I cannot find any relationship between those best sampling times and some characteristic time of the system.

The problem of multiple samples at multiple locations has been also considered for three simple cases. The sampling times are the same as in the previous case, that is, any time between 0 and T at intervals of 0.1T and the initial estimation covariance matrix is the diagonal one without spatial correlation. The first new case is the choice of two measurements at the location x=0.5T (55 possible sampling schemes) to see how a second measurement can improve the optimum result obtained with only one. The second case is the choice of two measurements at two different places x=0.5T and x=0.66T (232 possible sampling schemes) since it is not very reasonable to take a second measurement where the estimation error has been lowered by the first measurement. And the third case is again the choice of two samples

at two locations but in this case both locations are near one of the boundaries, x=0.16T and x=0.33T .

The optimal design for the first case can be obtained from Figure 5.14.

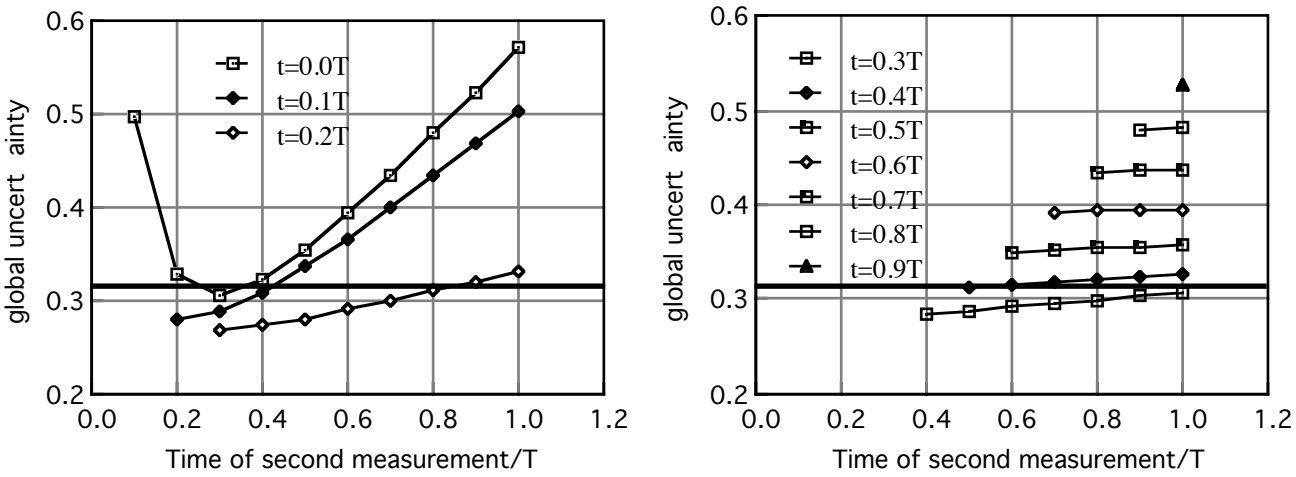


Figure 5.14. Mean estimation error when two measurements are taken at location x=0.5L. The time of the first measurement appears in the legend and the time of the second measurement is indicated in the x axis The horizontal bold line indicates the minimum value of uncertainty obtained with only one measurement.

There are 12 possible schemes that significantly improve upon the best result obtained with only one measurement and all of them include some measurement at times t=0.2T or t=0.3T, the best one being that which includes both of them. A reduction from 31% to 27% is obtained by the second measurement, which would probably not be justified by cost. Again, notice that there is something about t =0.3T that makes it optimum. The fact that the two measurements must be taken early in time in consecutive intervals is explained by the fact that it is more useful to take them during the transient state than in steady state, when a second measurement at the same place would be worth little.

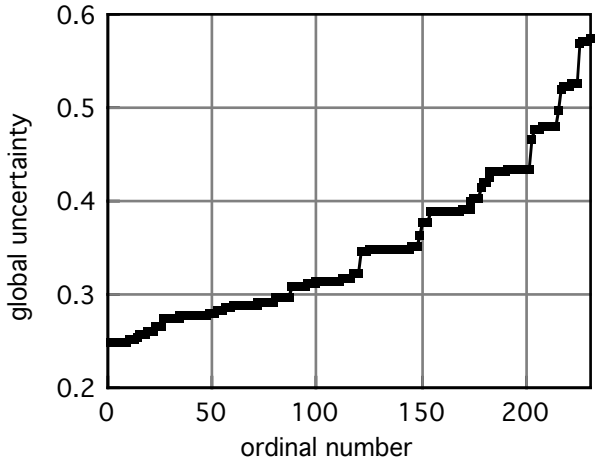


Figure 5.15. Ordered results of the 232 possible sampling schemes resulting from two samples at locations x=0.50L and/or x=0.66L

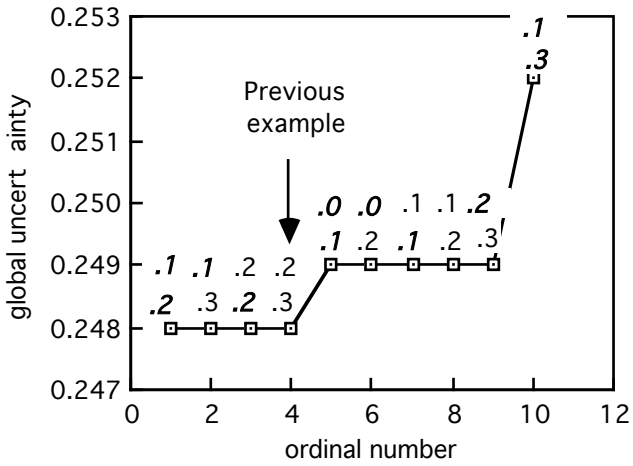


Figure 5.16. Blow-up of Figure 5.15 to show the best 10 possible sampling schemes. On top of each point it is written the time at which the sample must be taken. The location is represented by the typeface: plain face corresponds to x=0.5L and bold italic face to x=0.66L

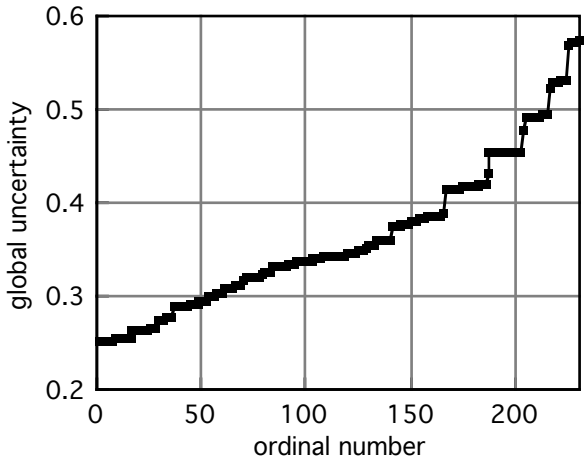


Figure 5.17. Ordered results of the 232 possible sampling schemes resulting from two samples at locations x=0.16L and/or x=0.33L

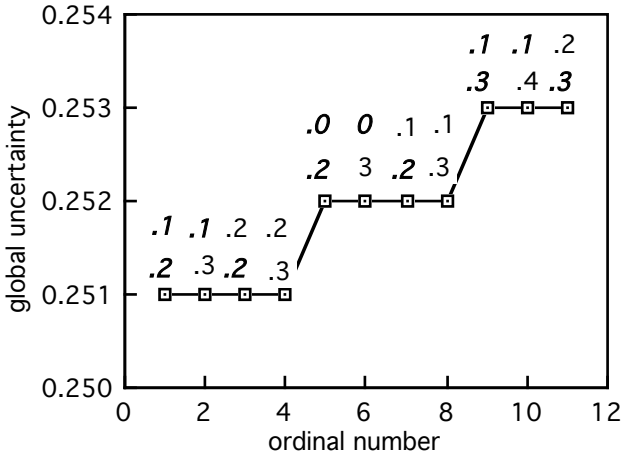


Figure 5.18. Blow-up of Figure 5.17 to show the best 10 possible sampling schemes. On top of each point it is written the time at which the sample must be taken. The location is represented by the typeface: plain face corresponds to x=0.16L and cursive bold face to x=0.33L

Figures 5.15 and 5.16 show the results for the second case. Three other possible sampling schemes give the same uncertainty as the optimum found in the previous case. About 90 schemes provides a mean error below 31% of the initial estimation error. The only conclusion that may be drawn from these results is that as the sampling moves closer to the boundaries the more effective samples are those taken at the beginning.

Figures 5.17 and 5.18 show that spatially uncentered samples do not give as good results as samples with one at the center. The number of sampling schemes that give errors below 31% are about 50 (half the number of the previous case). The optimum has a slightly larger uncertainty. It may be worthwhile to notice that in this case the four sampling schemes with the smallest uncertainty tend to sample early near the boundary (at x=0.16L) and later as the sample moves away from the boundary.

5.3. One-dimensional mass transport model

The same sort of network design problems addressed in the previous section will be presented here. The mechanics of the procedure are the same but the transition matrix changes substantially leading to completely different results.

The one-dimensional aquifer considered here has constant head gradient, constant hydraulic conductivity and constant dispersivity, which implies constant fluid velocity and constant dispersion coefficient. The aquifer is discretized in 30 segments of equal size and the simulation extends for a period equal to the mean advective travel time. The total simulation period is discretized into 1000 time steps. The presence of the advective term in equation (3.13) makes the solution of the problem more interesting and more difficult. On the one hand, the system has embedded in it an intrinsic time, which is related to the fluid velocity and the length of the aquifer, that could explain the timing of the samplings. On the other hand, numerical dispersion introduces a relatively large error in the state-transition model.

Both ends of the aquifer are constant concentration boundaries. The initial concentrations are zero all over the aquifer, except at x=0 where the concentration is one. The same comments made in the flow problem about the absence of the concentrations in the computation of the estimation covariance matrices can be made here, but in this case they will be presented together with the network design results because they give a feeling of how the travel time can determine the sampling timing. The boundary conditions are not very realistic for times larger than the breakthrough time since there is no sink for the solute that is constantly entering the system from the boundary at x=0.

The simplifications introduced for the one-dimensional aquifer and the choice of the implicit method to approximate the time derivative of concentration leads to the following transition matrix:

$$\mathbf{\Phi}=\begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ -\xi_1-\xi_2 & 1+2\xi_1 & -\xi_1+\xi_2 & 0 & \cdots & 0 \\ 0 & -\xi_1-\xi_2 & 1+2\xi_1 & -\xi_1+\xi_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \qquad \xi_1=\frac{D}{(\Delta x)^2}\Delta t \qquad \xi_2=\frac{v}{2(\Delta x)}\Delta t$$

where D is the dispersion coefficient (equal to the dispersivity times the velocity) and v is the fluid velocity (equal to the hydraulic conductivity times the hydraulic gradient over porosity). This transition matrix is no longer symmetric due to the fact that velocity has a clear directional component.

The state of the system as it is simulated by the model is shown in Figure 5.19, which gives a feeling for the timing of the advance of the front, the physical dispersion, and the numerical dispersion (the latter can be observed in the oscillations in the tail of the front as it advances).

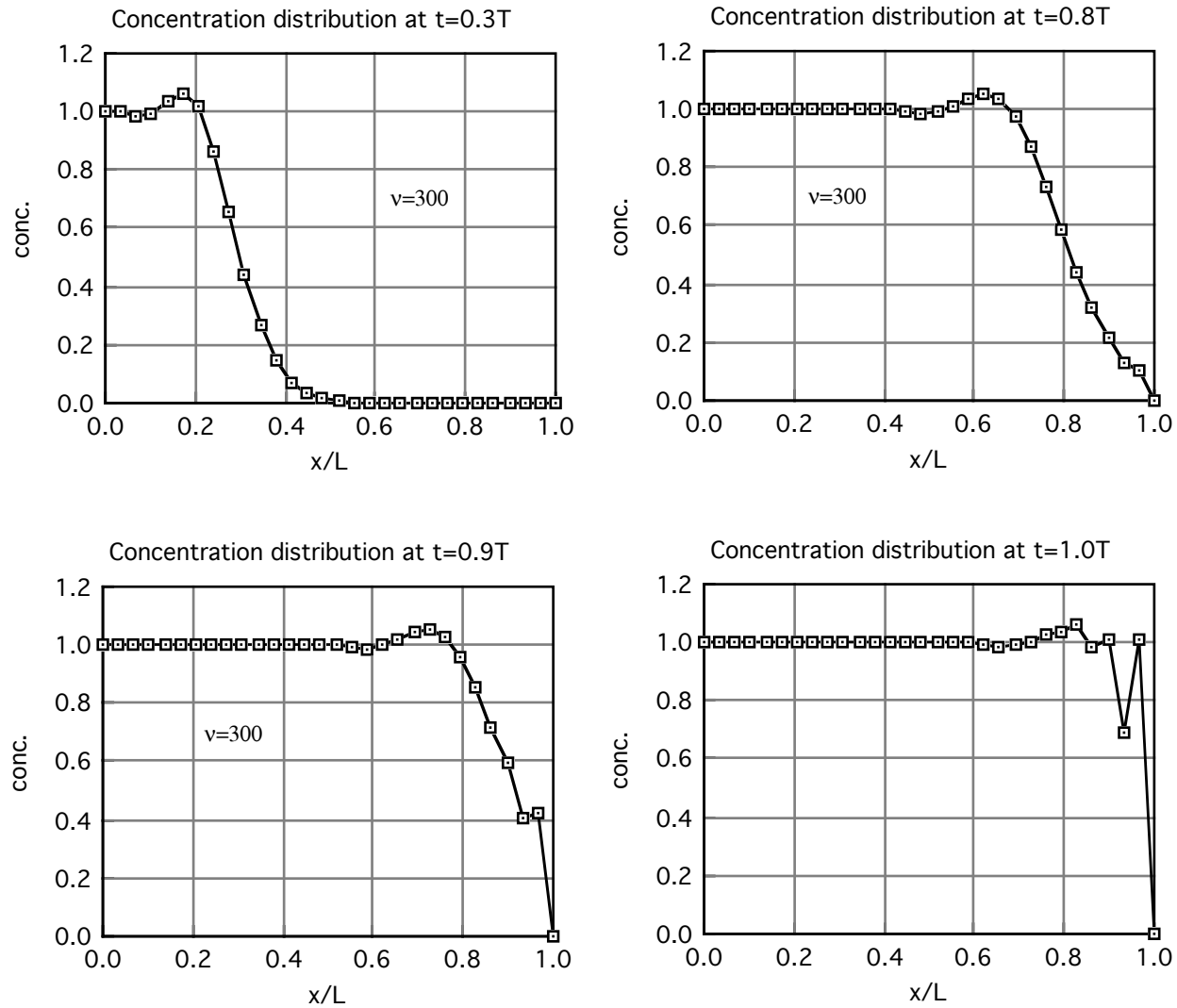


Figure 5.19. Four views of the state of the system at four different instants. The convective term is clearly dominant and the numerical dispersion is not very high until late in the simulation.

The initial value of the mean estimation error covariance matrix $\mathbf{\Sigma}(0|0)$ is again a diagonal matrix of the form $\sigma^2_0\mathbf{I}$ where $\mathbf{I}$ is the identity matrix and σ^2_0 is a large number. How this variance propagates in time, without considering sampling yet, is depicted in Figures 5.21—5.24 for the system I will call *basic*, which is characterized by ξ_1=0.3 and ξ_2=1.5 (Δt=1/10th of the travel time). One possible set of parameters that yields these values is v=1 m/d, Δx=10 m, Δt=30 d (total length L=300 m and travel time T=300 days), and longitudinal dispersivity, a_L=1 m (D=1 m^2/d). In terms of the Peclet number, the basic system is characterized by a Peclet number, ν=(vL/D)=300.

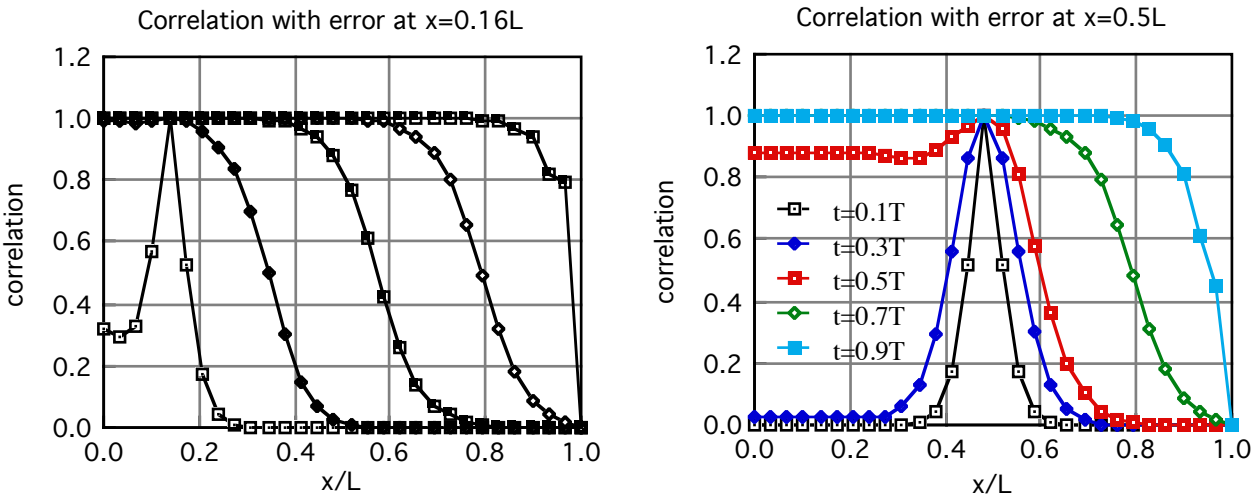


Figure 5.20. Evolution of the spatial correlation between errors at points x=0.16L and x=0.5L, and the estimation errors at the rest of the aquifer. (The legend applies to both plots)

The numerical model closely reproduces the behavior of the real system, with small numerical dispersion, except for times close to the travel time and for the locations next to the right end where the concentration is constant and equal to zero.

The evolution of the correlation between estimation errors has a clear physical significance (Figure 5.20). Initially the estimation errors are uncorrelated and at each point they remain nearly uncorrelated with the rest of the estimation errors until the front passes that point. From that moment on, the estimation errors at that point are fully correlated with the estimation errors of any "upstream" point and full correlation with "downstream" points keeps building up as the front moves on. The explanation is clear since each point remains at its initial state (no correlation) until the front reaches it. After a short time the concentration

becomes equal to the concentration at the left end and therefore fully correlated with this and with all the points in between. The behavior of the variance can be seen in Figures 5.21—5.23 and, as was the case for flow, is not as clearly explained.

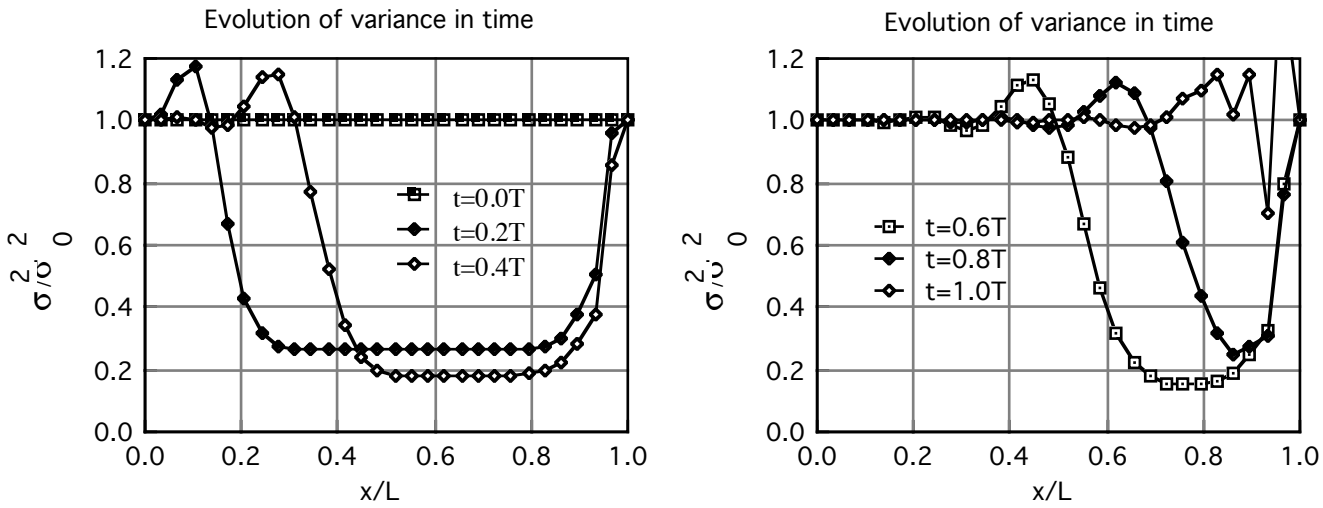


Figure 5.21. Variation in time of the estimation variance when no measurements are considered. Time and distance along the x-axis are relative to the total length and total travel time.

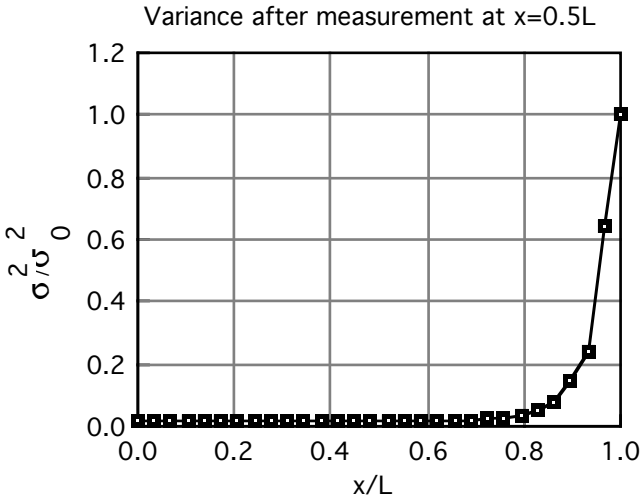


Figure 5.22. Estimation error variance at time= travel time after a measurement has been made at location x=0.5L.

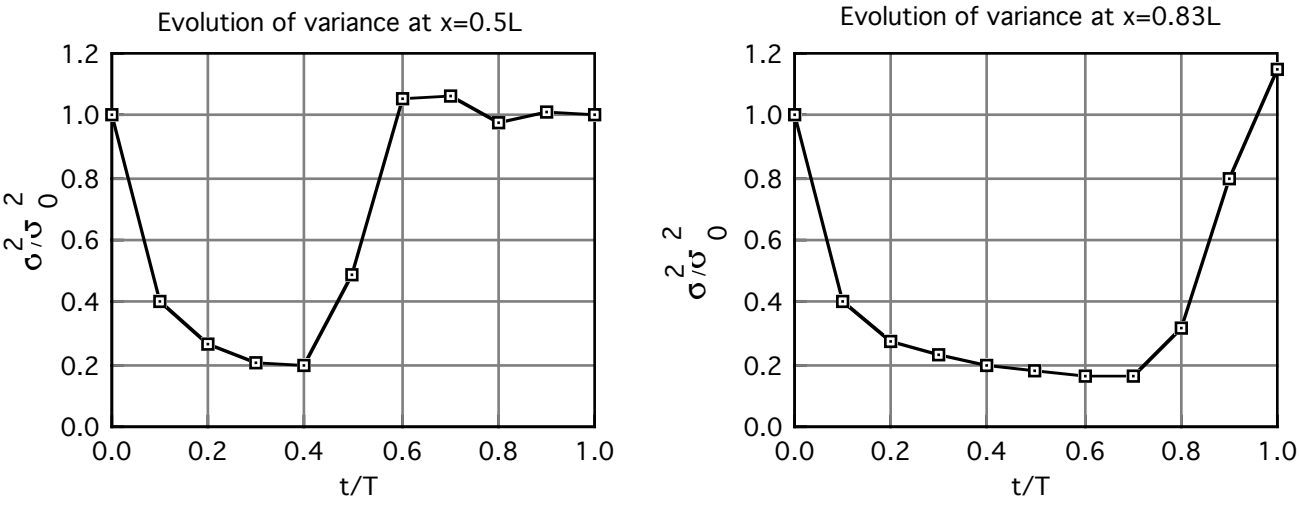


Figure 5.23. Evolution of estimation error variance at locations x=0.5L and x=0.83L

It is difficult to find a clear explanation why the estimation error variance decreases from its initial value to about 20% of its initial value until the front reaches the point. . The behavior before the front gets to the location means that our initial estimate of the state at that location gets better and better as time passes. To illustrate this, let us assume that the estimation errors are Gaussian so that the 95% confidence interval is four estimation standard deviations wide centered at the mean estimation. Figure 5.24 shows the evolution of the estimates predicted by the model and their corresponding confidence intervals. If the estimation covariance matrix at the end of the simulation period is filtered by an almost exact measurement at x=0.5L (in fact, the location does not matter) the variance reduces almost to zero.

The explanation for the reduction of variance shown in Figure 5.22 is clear since it was shown that total spatial correlation in the errors occurs after the front passes. Therefore, any measurement at time t=T will reduce the variance everywhere.

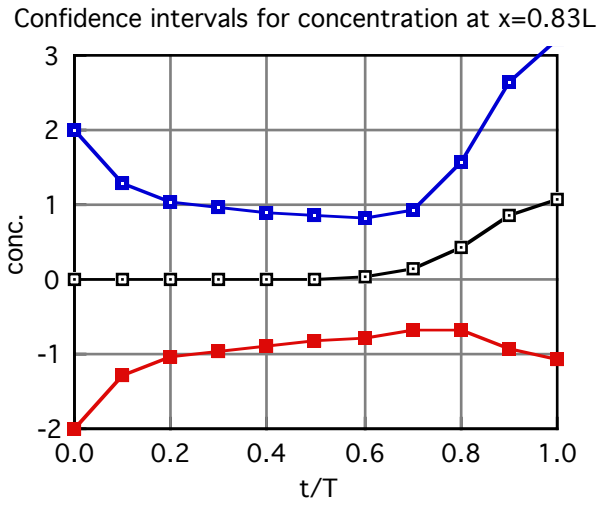


Figure 5.24. Assuming normality in the estimation errors the curves above represent the predicted concentration values at location 0.83L plus/minus two estimation error standard deviations. Notice how the confidence interval narrows around the estimates until the front reaches the point, then, the confidence interval widens.

The explanation for the behavior of the transient states must be looked for in the initial estimation covariance. The same network design problem was carried out for an initial estimation covariance with exponential spatial correlation of integral scale equal to one third of the aquifer length. The results are shown in Figure 5.25. This figure shows that when the initial estimation covariance matrix contains spatial correlation (in this case large enough to affect any two points and maintain uncorrelated the estimation errors at both ends) the evolution of estimation variance in time has a better physical explanation. The variance slightly decreases before the front gets to the point but remains around the same initial value and after the front has passed, the variance remains equal to the left end variance.

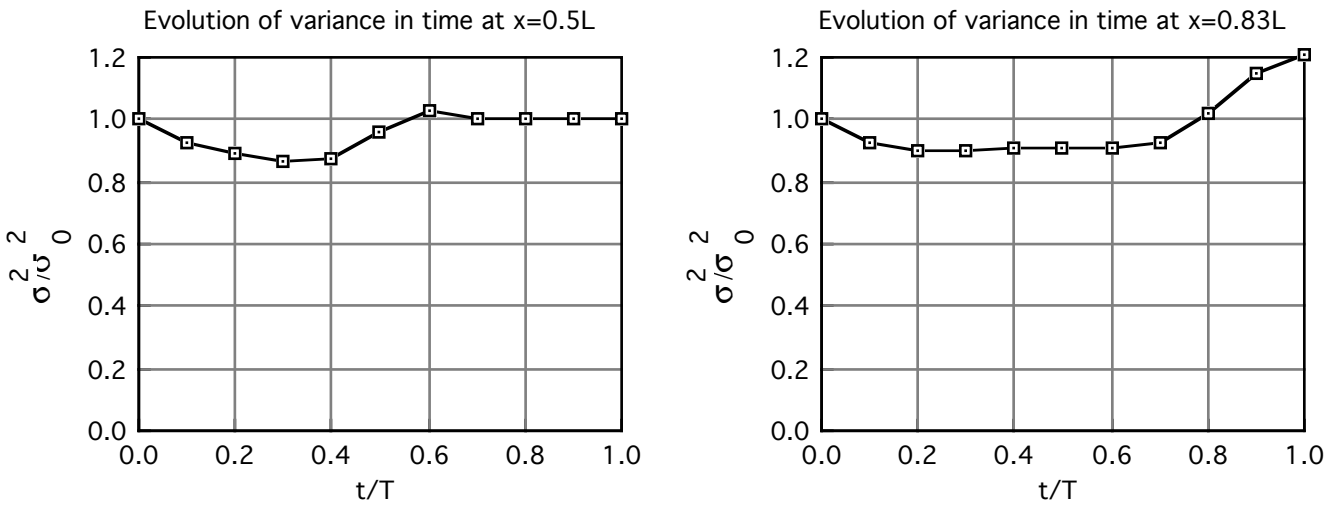


Figure 5.25. Propagation of variance in time at locations x=0.5L and x=0.83L when the initial estimation covariance matrix has an exponential correlation with integral scale equal to one third of the length of the aquifer.

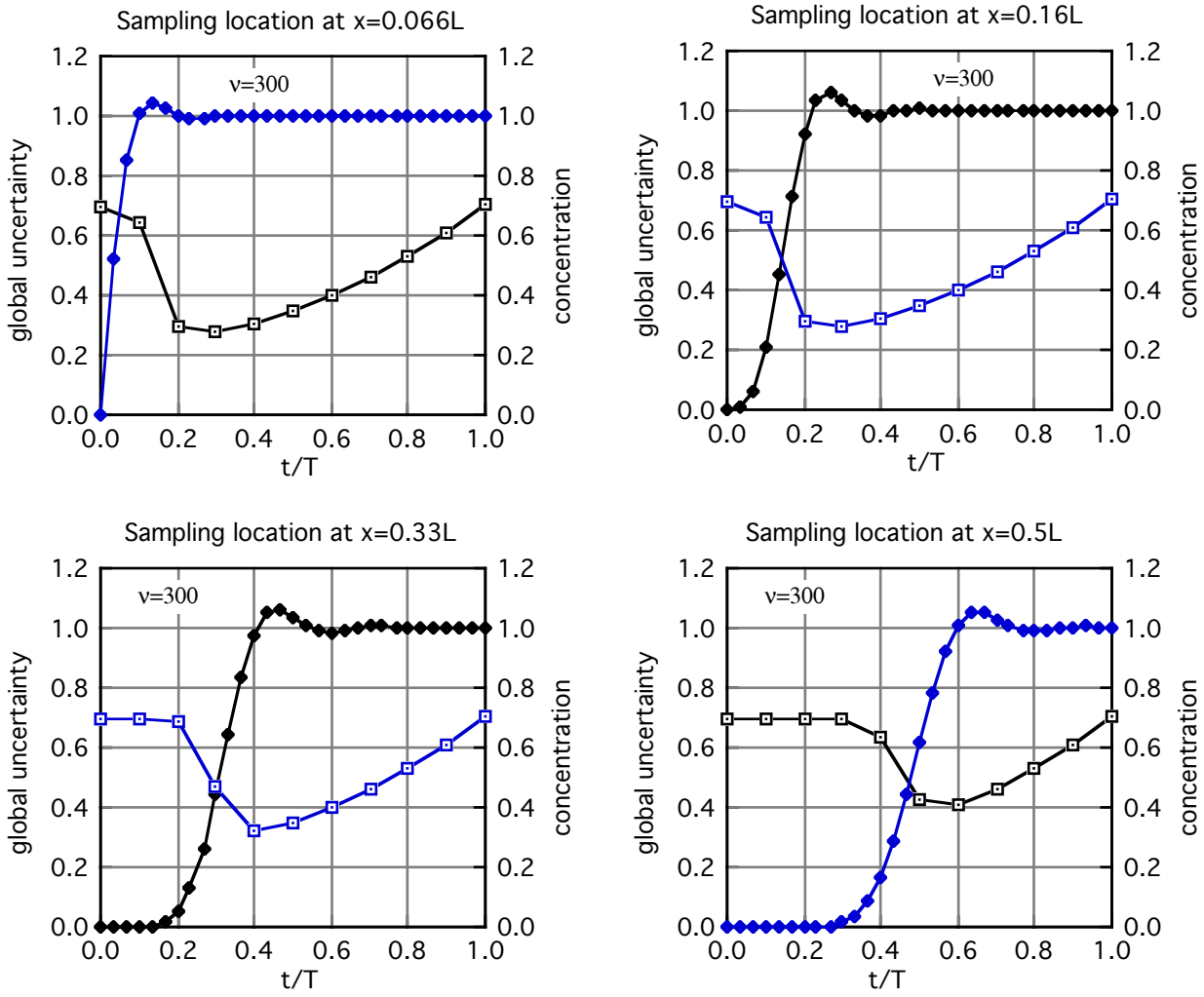
It seems that an initial exponential covariance gives results for the transient states with a heuristic explanation. This means that although we do not have any knowledge about the state of the system the fact that the physics of the phenomenon controlling the system behavior are known, is enough to introduce some correlation between intial errors.

A simple problem of network design is presented in the next paragraphs. The aquifer has the characteristics of the aforementioned *basic* problem. These can be summarized in a Peclet number of 300, and the initial estimation error covariance matrix is a diagonal matrix with a large constant variance at all points. The interest here will be to check the effect of the Peclet number. Notice that the *basic* problem has been chosen with a large Peclet number, so that the convective term is highly dominant over the dispersive term Low Peclet numbers would provide the same results as for the previously discussed flow problem, which is equivalent to a completely dispersive mass transport problem.

The first case is to study the selection of the single point in time and space that provides the sample that most reduces the global uncertainty. The search is constrained to the same set of 77 points defined by the cartesian product (x ,t)={0.066L, 0.16L, 0.33L, 0.5L, 0.66L, 0.83L, 0.93L }×{0.0T, 0.1T, 0.2T, 0.3T, 0.4T, 0.5T, 0.6T, 0.7T, 0.8T, 0.9T }. In all these results the breakthrough curve simulated by the numerical model from initial conditions corresponding to zero concentration everywhere except at the left end is plotted along with the uncertainty, to

see how the travel time affects the possible choices. At the risk of being repetitive, let me reiterate that the actual values of the initial concentrations or the concentrations at any time have no effect at all on the computation of the estimation covariance matrix and that the results would have been the same for any other initial concentration distribution as long as the initial covariance matrix and the boundary conditions are the same. But, in the case of mass transport, the travel time is built in to the transfer matrix $\mathbf{\Phi}$(through the velocity and the discretization of x). Therefore, the concentrations simulated by the model, which correspond to quite a sharp front moving from left to right, will reflect how the travel time affects the solutions.

Figure 5.26 shows, for each sampling point, the global uncertainty associated with each sampling time as well as the breakthrough curve at that point.



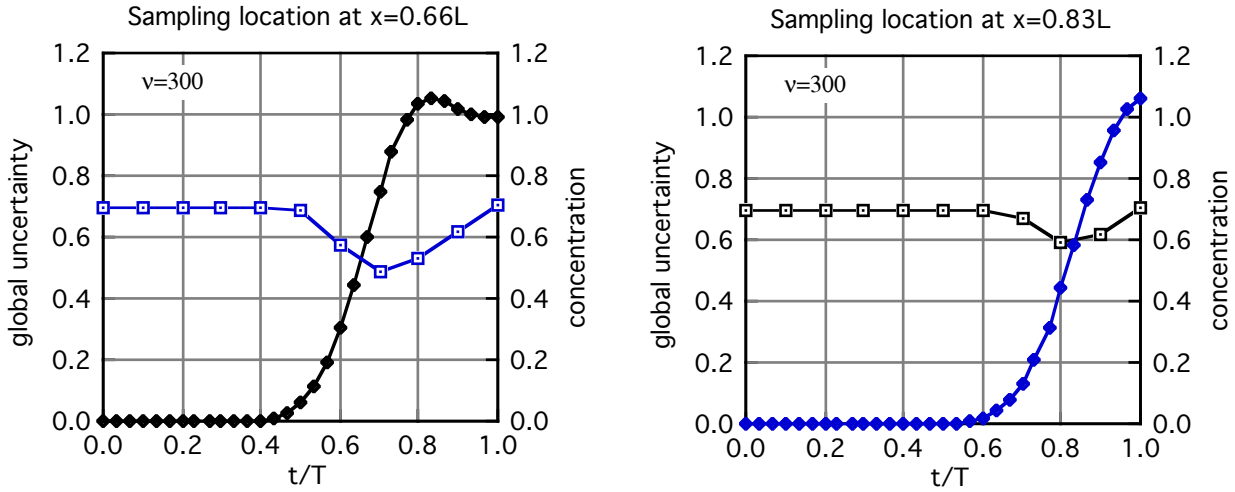


Figure 5.26. Continued on next page.

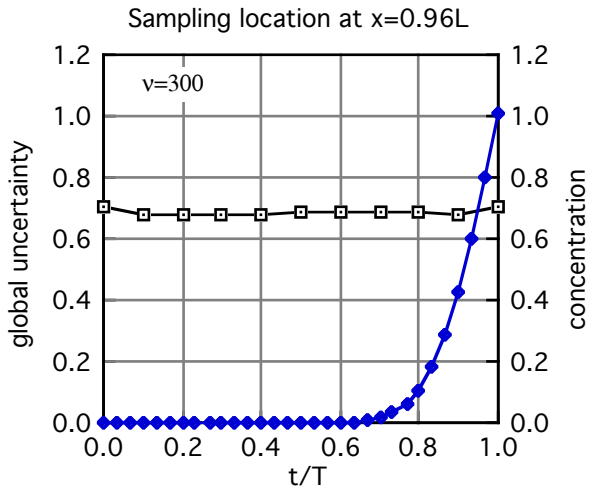


Figure 5.26. Each curve shows: in white, the mean uncertainty over time and space if a sample had been taken at the time and location specified; in black the breakthrough curve at the location indicated.

From these Figures a number of comments can be made. It seems that there is an uncertainty value around 0.7 that is filtered by the model no matter when or where the sample is done. This can only be explained in terms of the initial estimation error covariance matrix and corresponds to the decrease of the variance at any point in the first time steps of the simulation (see Figures 5.21—5.24). The question that remains unanswered is the reason for that decrease. A sort of bulge appears below the 0.7 uncertainty line that moves from left to right as the sampling location does. The minimum uncertainty happens shortly after or at the breakthrough time and it is smaller the closer to the left end the sampling point is. That is, the sooner the front is detected the better.

From Figure 5.27 is clear that the best sampling time increases as the sampling point approaches the right end and that the reduction in uncertainty is larger if the sample is taken near the left end.

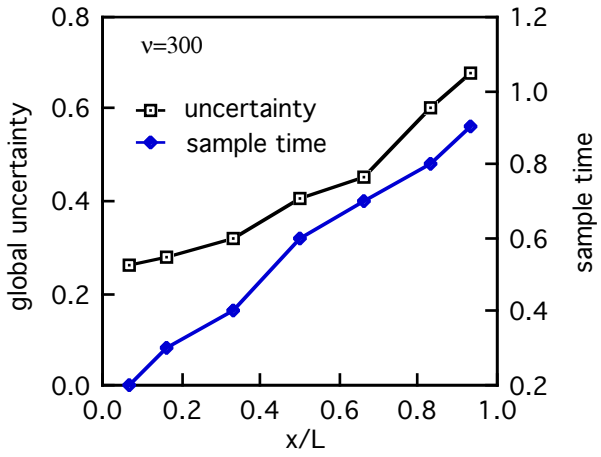


Figure 5.27. For each sampling point the best time to make the sample and its associated uncertainty are displayed.

Due to the behavior observed for the variance propagation for uncorrelated initial errors the same exercise was done in the case of correlated initial errors with an exponential covariance of integral scale equal to one third of the length of the aquifer. The results are shown in Figures 5.28—5.29.

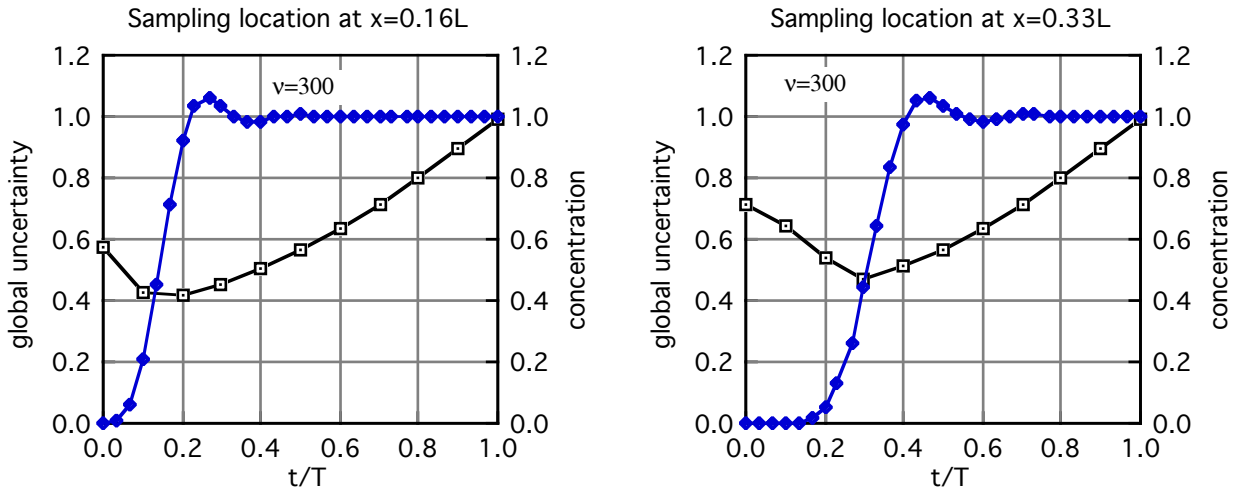


Figure 5.28. Continued on next page.

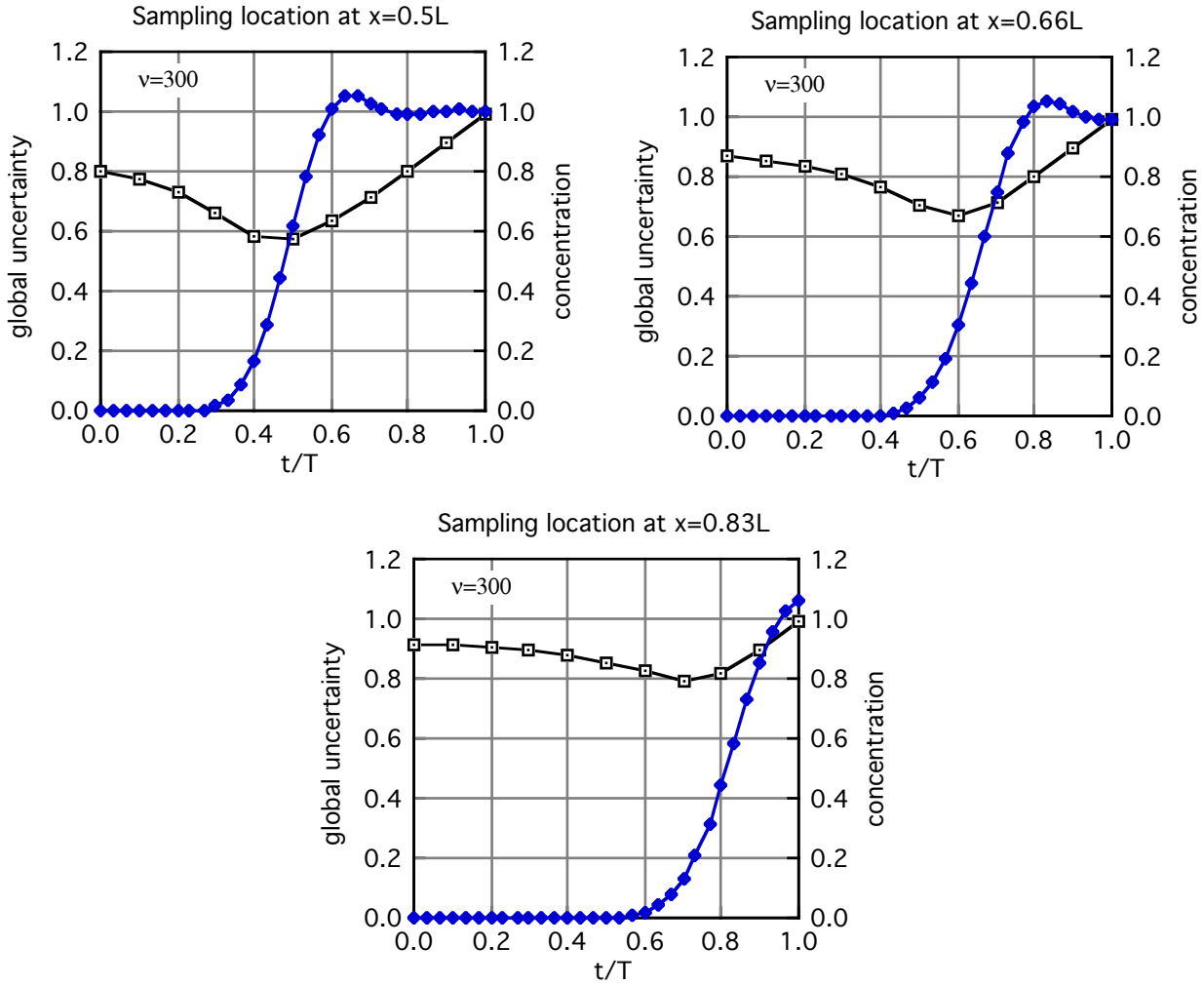


Figure 5.28. Each curve shows: in white, the mean uncertainty over time and space if a sample had been taken at the time and location specified; in black the breakthrough curve at the location indicated. An initial exponential correlation in the errors has been included.

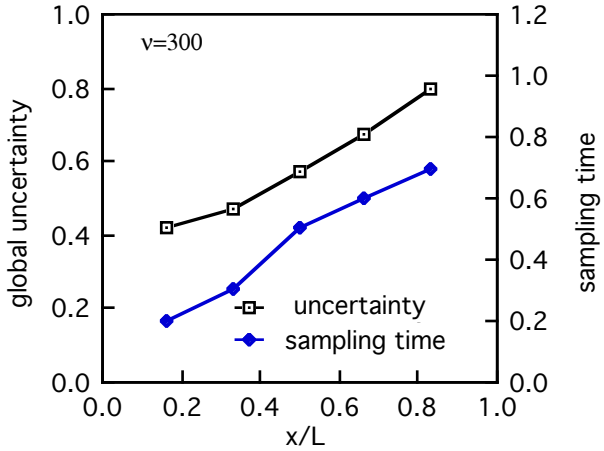


Figure 5.29. For each sampling point in Figure 5.28, the best time to take the sample and its associated uncertainty are displayed.

Some changes can be observed from the results without initial correlation. The sill that appeared at about 0.7 moves to 0.95, meaning that in this case the model is not able to filter as

much uncertainty without the help of some measurements. The uncertainty values are larger but the sample locations and timing are almost the same. The optimum is achieved early in time and close to the left end.

A second exercise will take a look at the effect of the Peclet number in the selection of the best time to take a sample at x=0.5L. The uncorrelated initial estimation covariance matrix is used. The Peclet number will take the values 30, 75, 150, 300, 600, 1200, 3000 and 30000. These values are obtained by changing the longitudinal dispersivity and maintaining the gradient. A change in the gradient would have affect both velocity and dispersion so that the Peclet number would have remained unchanged and so would the transfer matrix. The breakthrough curves are displayed to show the amount of physical dispersion and the effect of numerical dispersion. The results are displayed in Figures 5.30 and 5.31.

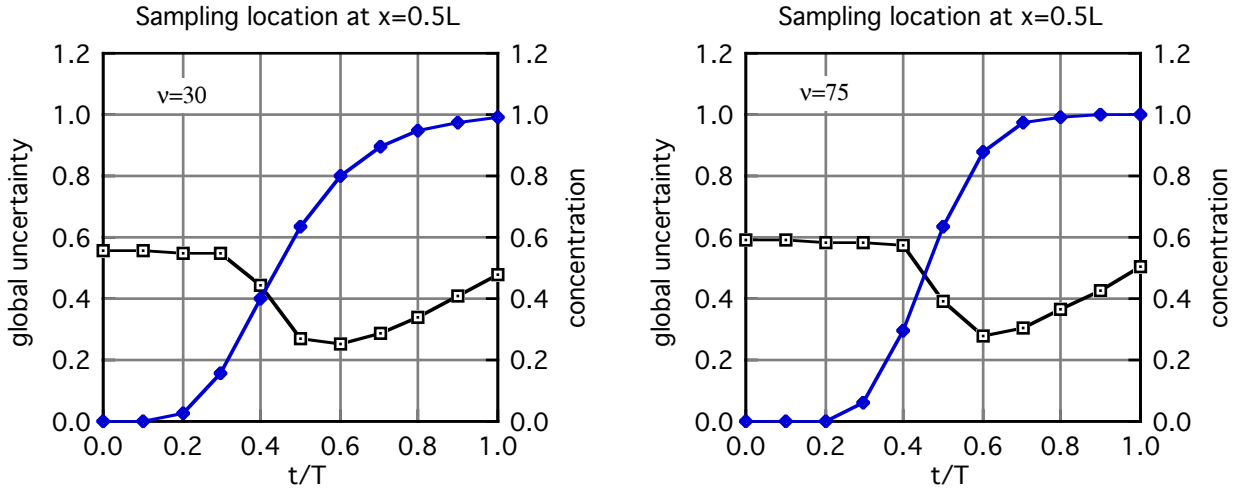


Figure 5.30. (Continued on next page).

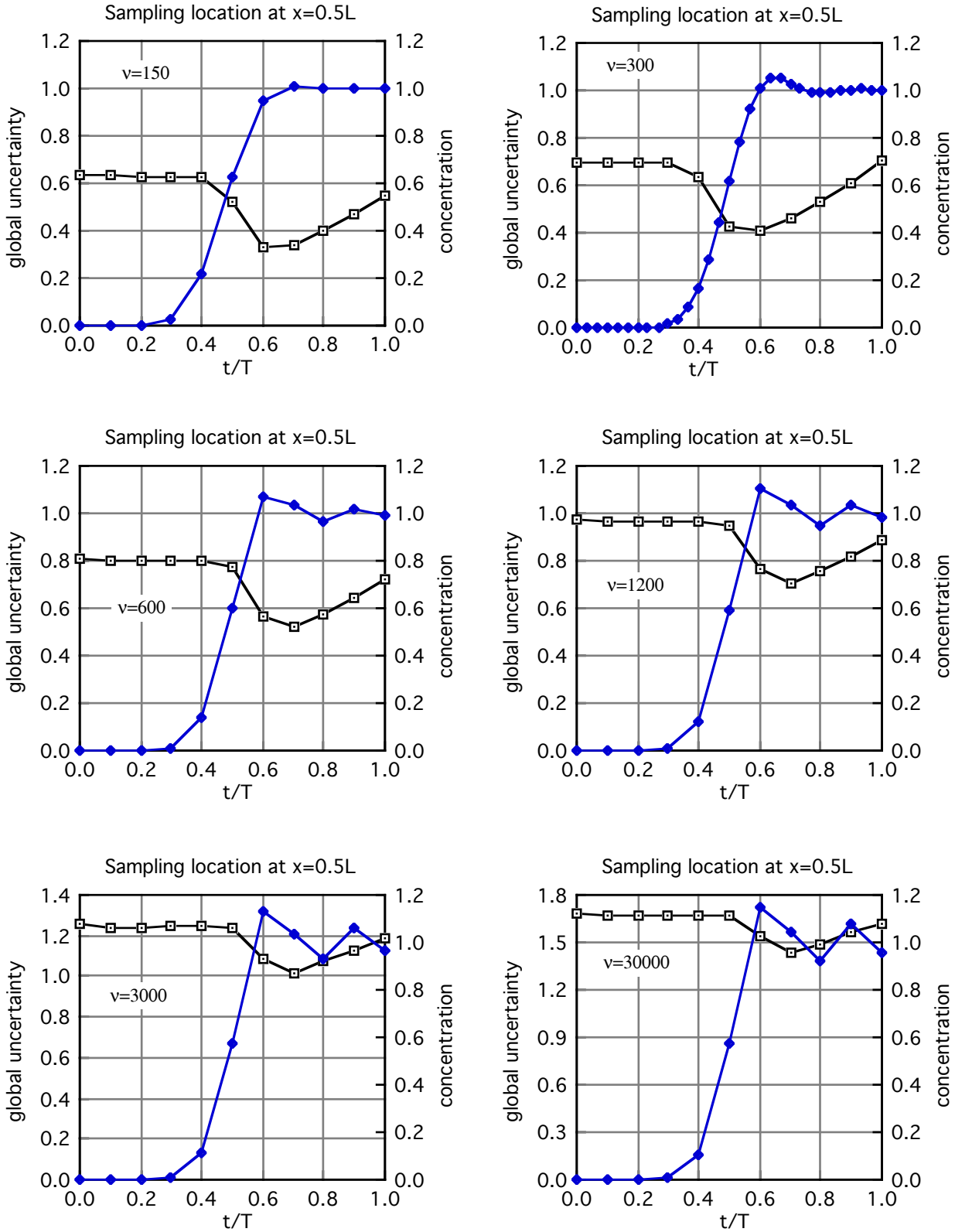


Figure 5.30. Each curve shows, for several Peclet numbers: in white, the mean uncertainty over time and space if a sample had been taken at location x=0.5L and at the time indicated; in black the breakthrough curve.

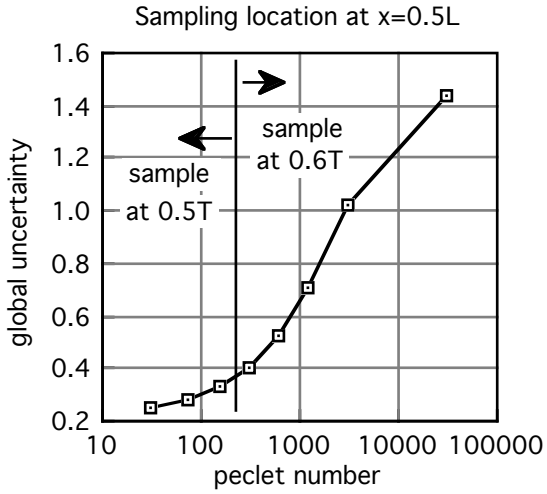


Figure 5.31. Uncertainty corresponding to the best sample in time at location x=0.5L, for several Peclet numbers. The time was either 0.5T or 0.6T as shown in the graphic.

The results for the largest values of the Peclet number should be discarded due to the large numerical dispersion presented by the model. The breakthrough curves show large oscillations after the front has reached x=0.5L. Two important conclusions can be drawn from these Figures: i) the timing of the samples does not seem to depend on the Peclet number but rather on the travel time and ii) the dispersion (or the Peclet number) plays an important role in the level of uncertainty that can be removed from the system by sampling, and it is clear that the larger the dispersion (smaller Peclet number) the better the system can be known. The explanation can be seen in the fact that in a pure convective system a measurement only can give information about the system in one direction, "upstream" if the front already passed the sampling point, or "downstream" if the front is yet to pass, whereas in a pure dispersive system the information obtained at a point will travel in both directions.

Some other results that are not plotted here are the following: A case of multiple measurement at multiple locations was studied. The improvement due to the second sample was small and to my understanding the decrease in uncertainty would not justify the cost of it. The timing for the best scheme in the two cases studied was the union of the timings for the best sample when each point was considered alone.

The only effect of big measurement errors (50% of the initial variance) in the solution of the network design problem was an increase in uncertainty but no effect on the location or timing of the optimal sample.

6. CONCLUSIONS AND FUTURE RESEARCH

It is clear that this study is limited in its scope and that a large number of avenues of research remain open. But, in any case, I would like to point out the conclusions to which this work has led me and some goals for future research.

The Kalman filter is a tool suitable to be used in network design for any dynamic phenomenon that can be represented by a linear state-transition model and for which an estimate of the initial system as well as the error covariance matrix of that estimate are known. Notice that no assumption of normality or of any other probabilistic kind was needed to develop the method. The only two requirements are the knowledge of the initial estimate and its estimation variance.

The choice of the initial estimation error covariance matrix is of major influence in the way variance is propagated. In particular, the differences between starting with and without spatial correlation were studied and found significant, especially in reference to the final value of the global uncertainty, although not so much in the network design results.

An optimal design in the way done here, i.e., combinatorial search, is infeasible for practical purposes due to the dimensions of the problem. The discretization in time is what increases very much the dimensions. If the time for the samples were fixed the selection of the best sampling locations at given times would be much faster. Together with the dimensionality problem, the mass transport case requires a very fine time discretization in order to obtain accurate results using a finite difference approach.

In the flow problem, the location of the samples depends mainly on the geometry of the system and on the initial estimation error covariance. The timing of the samples is more difficult to relate to a property of the system.

In the mass transport problem, the location of the samples seems to depend on the boundary and initial conditions whereas the timing depends on the travel time. The only influence of the Peclet number is on the amount of uncertainty but not on the network design itself.

I find three important weak points in this work that must be carefully studied.

First, the way the optimization has been approached. A better way to perform it must be found. Branch and bound methods could be good candidates if some algorithm to discard schemes before computing them can be found.

Second, the fact that the design procedure is independent of the data values themselves. I have emphasized this point throughout this work, especially when actual head or concentration values were displayed for the sake of clarity, I would like to point out that this property, which has been regarded by many researchers as an advantage since it allows the posterior estimation variance to be known prior to the actual measurement, is a weak point. The information provided by the data as it is collected must be used not only to predict the state of the system but also to asses its estimation error. Nonlinear filters, such as disjunctive kriging or indicator kriging, that provide data- conditioned estimation variances could be the tool to use.

And third, parameter uncertainty was not considered. All the parameters defining the system, as well as the external actions, were considered exactly known. But parameters vary in such a way that even in the event of perfect measurements this variability could not be modeled from a few samples in the field, making it impossible to know them with accuracy. The introduction of parameter uncertainty is the first big step that must be done in the continuation of this work and it should be included in such a way that the objective of the network design problem is not only to minimize the uncertainly about the system but also the uncertainty about the parameters.

And finally, other points that require further research are: the objective function (in this case, minimization of the uncertainty about the system) should be written in terms of the eventual decision variables (risk of contamination, amount of insurance to buy, etc.) including economic costs, social costs and anything that relates to the effect of the system uncertainty on the final decision making-process; the model error, i.e., the error coming from a model that is a simplification of reality, an error that is on top of the error induced by parameter uncertainty, should be considered, and, of course, some real two-dimensional cases should be studied.

References

Bastin, G, B. Lorent, C. Duqué and M. Gevers, (1984), *Optimal Estimation of the Average Areal Rainfall an Optimal Selection of Rain Gauge Locations*, Water Resources Research, 20(4), 463—470.

Bear, J., (1979), *Hydraulics of Groundwater*, McGraw—Hill, 569 pp., New York.

Bogárdi, I and L. Duckstein, (1985), *Multicriterion Network Design Using Geostatistics*, Water Resources Research, 21(2), 199—208.

Bracewell, R., (1978), *The Fourier Transform and Its Applications,* McGraw-Hill, 445 pp., New York.

Bras, R. and I. Rodríguez Iturbe, (1985), *Random Functions and Hydrology*, Addison—Wesley Publishing Company, 557 pp., Massachusetts.

Bras, R. and I. Rodríguez Iturbe, (1976a), *Network Design for the Estimation of the Areal Mean of a Rainfall Event by a discrete summation*, Water Resources Research, 12(6), 1185—1195.

Bras, R. and I. Rodríguez Iturbe, (1976b), *Rainfall network design for runoff prediction*, Water Resources Research, 12(6), 1197—1208.

Bras, R. and R. Colón, (1978), *Time Averaged Areal Mean of Precipitation: Estimation and Network Design*, Water Resources Research.

Carrera, J., Usunoff, E and F. Szidarovszky, (1984), *A Method for Optimal Observation Network Design for Groundwater Management*, Journal of Hydrology, 73, 147—163.

Caselton, W. and T. Husain, (1980), *Hydrologic Networks: Information Transmission*, Journal of the Water Resources Planning and Management Division, 106(WR2), 503—520.

Chou, D., (1986), *Model and Algorithm of Selecting Optimum Exploratory Drill Locations by Group*, Submitted to Mathematical Geology.

Dandy, G.C., (1976), *Design of water quality sampling systems for river networks*, Ph. D. thesis, Massachusets Institute of Technology, Cambridge.

Delfiner, P., (1975), *Linear estimation of nonstationary spatial phenomena, in*, Advanced Geostatistics in the Mining Industry, edited by M. Guarascio, pp. 49—68, D. Reidel, Dordrecht, Holland, 1975 .

Delhomme, J.P., (1978), *Kriging in the Hydrosciences*, Advances in Water Resources, 1(5), 251—266.

Eagleson, P.S. and M.J. Goodspeed, (1973), *Linear systems techniques applied to hydrologic data analysis and instrument evaluation*, Technical Report 34, Commonwealth Scientific and Industrial Research Organization Division of Land Use, Melbourne, Australia .

Eagleson, P.S., (1967), *Optimum Design of rainfall networks*, Water Resources Research, 3(4), 1021.

Golub, G.H., (1983), *Matrix Computations*, Johns Hopkins University Press, 476 pp., Baltimore.

Grayman, W.N. and P.S. Eagleson, (1971), *Evaluation of radar and rain gage systems for flood forecasting*, Technical Report 138, Ralph M. Parsons Laboratory for Water Resources and Hydrodynamics, Massachusetts Institute of Technology, Cambridge.

Hoel, P.G., S.C Port and C.S. Stone, (1972), *Introduction to Stochastic Systems*, Houhgton Mifflin, 203 pp., Boston.

Hughes, J.P. and D.P. Lettenmaier, (1981), *Data Requirements for Kriging: Estimation and Network Design*, Water Resources Research, 17(6), 1641—1650.

Jones, D.A, Gurney R.J. and P.E. O'conell, (1979), *Network Design Using Optimal Estimation Procedures*, Water Resources Research, 15(6), 1806—1812.

Journel, A.G. and C.J. Huijbregts, (1978), *Mining Geostatistics*, Academic Press, 600 pp., New York.

Kalman, R.E., (1960), *A New Approach to Linear Filtering and Prediction Problems*, Journal of Basic Engineering, Transactions of the ASME, pp. 35—45.

Kalman,. R.E. and R.S. Bucy, (1961), *New Results in Linear Filtering and Prediction Theory*, Journal of Basic Engineering, Transactions of the ASME, pp. 95—108.

Kitanidis, P., C.S Queiroz and D. Veneziano, (1978), *Sampling networks for violation of quality standards*, in Proceeding of AGU Chapman Conference on Kalman filters, edited by C.L. Chiu, pp. 213—230 , Stochastic Hydraulics Program, Department of Civil Engineering, University of Pittsburgh, Pittsburgh, PA.

Lapidus, L. and G.F. Pinder, (1982), *Numerical Solutions of Partial Differential Equations in Science and Engineering*, Wiley—Interscience, 677 pp, New York.

Lettenmaier, D. (1975), *Design of monitoring systems for detection of trends in stream quality*, Technical Report 39, C.W. Harrus Hydraulic, Laboratory, Department of Civil Engineering, University of Washington, Seattle.

Lettenmaier, D., (1979), *Dimensionality Problems in Water Quality Network Design*, Water Resources Research, 15(6), 1692—1700.

Liebelt, P.B., (1967), *An Introduction to Optimal Estimation*, Addison-Wesley Publishing Company, 273 pp, Massachusetts.

Matheron, G., (1971), *The theory of regionalized variables and its Applications*, Cahiers du Centre de Morphologie Mathématique, 5, 211 pp, Ecole des Mines, Fontainebleau.

Matheron, G., (1973), *The intrinsic random functions and its applications*, Advances in Applied Probability, 5, 438—468.

McBratney, A.B., Webster, R. and T.M. Burgess, (1981a), *The Design of Optimal Sampling Schemes for Local Estimation and Mapping of Regionalized Variables—I*, Computers and Geosciences, 7(4), 331—334.

McBratney, A.B., Webster, R. and T.M. Burgess, (1981b), *The Design of Optimal Sampling Schemes for Local Estimation and Mapping of Regionalized Variables—II*, Computers and Geosciences, 7(4), 335—365.

McLaughlin. D., (1978), *Potential Applications of Kalman Filter Concepts to Groundwater Basin Management*, in Proceeding of AGU Chapman Conference on Kalman filters, edited by C.L. Chiu, pp. 639—658, Stochastic Hydraulics Program, Department of Civil Engineering, University of Pittsburgh, Pittsburgh, PA.

Moore, S.F., (1971), *The applications of linear filter theory to the design and improvement of measurement systems for aquatic environments*, Ph. D. Thesis, University of California at Davis.

Moore, S.F., (1973), *Estimation theory applications to design of water quality monitoring systems*, Journal of Hydraulic Division of the American Society of Civil Engineering, 99(HY5), 815—831.

Papoulis, A., (1984), *Probability, Random Variables and Stochastic Processes*, McGraw—Hill, 576 pp., New York.

Pimentel, K.D., (1975), *Toward a mathematical theory of environmental monitoring: The infrequent sampling problem*, Ph. D. thesis, University of California at Davis.

Pimentel, K.D., (1978), *Asymptotic error growth applied to monitoring*, in Proceeding of AGU Chapman Conference on Kalman filters, edited by C.L. Chiu, pp. 213—230 , Stochastic Hydraulics Program, Department of Civil Engineering, University of Pittsburgh, Pittsburgh, PA.

Pinder, G.F and Gray, (1977), *Finite Element Simulation in Surface and Subsurface Hydrology*, Academic Press, 295 pp., New York.

Remson, I., G. Harnberger and F. Molz, (1971), *Numerical methods in subsurface hydrology with an introduction to the finite elements method*, New York Wiley Interscience.

Rodríguez Iturbe, I and J.M. Mejía, (1974), *On the Design of Rainfall Networks in Time and Space*, Water Resources Research, 10(5).

Sanders, T.G. and D.D Adrian, (1978), *Sampling Frequency for River Quality Monitoring*, Water Resources Research, 14(4), 569—576.

Schweppe, F.C, (1973), *Uncertain Dynamic Systems*, Prentice—Hall International, 563 pp., Englewood Cliffs, New Jersey .

Sophocleous, M., (1983), *Groundwater Observation Network Design for the Kansas Groundwater Management Districts, USA*, Journal of Hydrology, 61, 371—389.

Sophocleous, M., Paschetto, J. E. and R.A. Olea, (1982), *Groundwater Network Design for Northwest Kansas Using the Theory of Regionalized Variables*, Groundwater, 20(1), 48—58.

Switzer, P., (1979), *Statistical Considerations in Network Design*, Water Resources Research, 15(6), 1712—1716.

Trescott, P.C., G.F Pinder and S.P. Larson, (1976), *Finite—Difference Model for Aquifer Simulations in Two Dimensions with Results of Numerical Experiments*, Techniques of Water—Resources Investigations of the United States Geological Survey, United State Government Printing Office, Washington.

Villeneuve, J.P., G. Morin, B. Bobee, D. Leblanc and J.P. Delhomme, (1979), *Kriging in the Design of Streamflow Sampling Networks*, Water Resources Research, 15(6), 1833—1840.

APPENDIX A. MATRIX INVERSION LEMMA.

We want to prove the following inversion identity:

$$(I_n + AB)^{-1} = I_n - A(I_m + BA)^{-1}B$$

where

A is an arbitrary $n \times m$ matrix

B is an arbitrary $m \times n$ matrix

$\mathbf{I}_n$ is the $n \times n$ identity matrix

$\mathbf{I}_m$ is the $m \times m$ identity matrix

If ($\mathbf{I}_m + \mathbf{AB}$) is not singular, then

$$-AB = -A(I_m + BA)(I_m + BA)^{-1}B$$

adding ($\mathbf{I}_n + \mathbf{AB}$) to both sides,

$$I_n = (I_n + AB) - A(I_m + BA)(I_m + BA)^{-1}B$$

or,

$$I_n = (I_n + AB) - (A + ABA)(I_m + BA)^{-1}B$$

or,

$$I_n = (I_n + AB) - (I_n + AB)\,A\,(I_m + BA)^{-1}B$$

or,

$$I_n = (I_n + AB)(I_n - A\,(I_m + BA)^{-1}B)$$

then, if $\mathbf{I}_n + \mathbf{AB}$ is not singular, premultiplication of both sides by $(\mathbf{I}_n + \mathbf{AB})^{-1}$ completes the proof.

The significance of this matrix inversion identity is that (if $n > m$) then the inverse of a large matrix can be obtained from the inversion of a smaller one.