Uncertainty Introduced by Upscaling Type Transfer Functions

Iris T. Stewart and Keith Loague
Department of Geological and Environmental Sciences
Stanford University
Stanford, CA, 94305-2115, USA
Ph. +1 650-723-4754; Fax +1 650-725-0979
Email: iris@pangea.stanford.edu

Keywords: soil science, model error, spatio-temporal uncertainty, stochastic simulation, sensitivity analysis

Abstract
The objective of the effort reported here was to evaluate the uncertainty resulting from upscaling type transfer functions (TTFs), which are used for quantitative, regional-scale agrochemical leaching estimates. TTFs are representative transfer functions for each soil textural class that yield concentration breakthrough time series for a depth of interest for ranges of conditions relevant to the transport processes that are important at the regional scale. This study focused on the Loam textural class and on four selected sets of TTFs. Each selected set of TTFs represents an increasingly greater degree of upscaling, with the total number of TTFs in a set ranging from eight to one. TTF performance was compared to blocks sampled from a field of process-based simulated stochastic concentration time series for the Loam soil texture. Each set of TTF is tied to saturated hydraulic conductivity and soil moisture information and applied to each block for the three sampling strategies. The results show that uncertainty in TTF concentration estimates is significantly reduced when using the set containing two TTFs as compared to the upscaled set containing one TTF. The uncertainty in TTF estimates was not consistently reduced further by adding more TTFs to the set.

1. Introduction
Large-scale, modern agricultural practices involve the application of agriculturally related chemicals, such as pesticides and fertilizers over areas spanning hundreds of square kilometers. Depending on the depth of the water table, the chemical applied, the soil hydraulic parameters, and the climatic conditions, agrochemical pollutants can potentially leach to the water table and threaten water resources (*Corwin et al., 1999*). In the framework of regulatory decisions relating to agricultural policies and water resource protection it is of fundamental importance to establish quantitative groundwater vulnerability estimates of pesticide leaching at regional scales. For example, given information on a pesticide application at the surface, the type of soils present, and the climatic conditions for an area under investigation, regulators need the capability to produce a map or GIS database of solute concentrations in space and time. However, these quantitative estimates are difficult or impossible to obtain, as both regional-scale, unsaturated zone solute concentration measurements, and regional-scale process-based numerical modeling have proven to be too expensive and time consuming (*Corwin et al., 1999*).

Recently, the idea of upscaling a linear systems or transfer function (TF) approach, such that it can provide quantitative estimates of groundwater vulnerability on the regional scale, has been brought forward (*Stewart and Loague, 1999*). Transfer functions are impulse response functions that characterize the internal dynamics of a system implicitly through a probability density function. The new regional-scale approach is founded on the work of a previous generation of researchers (e.g., *Jury, 1982; Jury and Roth, 1990; Jury et al., 1982; White et* *al., 1986*) that established the TF approach as a viable alternative in field or plot scale vadose zone contaminant leaching problems. The Jury Convective Lognormal Transfer function model (Jury CLT) is now well recognized. The work of *Stewart and Loague* (*1999*) has extended the Jury CLT to the regional scale through the use of type transfer functions (TTFs) in the context of non-point source agrochemical leaching assessments. TTFs are representative transfer functions, that allow the estimation of solute concentrations in space and time, not only at one location with one set of soil hydraulic parameters, but for predefined ranges of conditions relevant to the transport processes that are important at the regional scale. Once the best set of TTFs is determined, concentration estimates can be obtained by appropriately scaling a known function. Therefore, TTFs represent an alternative, cost-effective method to quantitatively assess concentrations in space and time for regional-scale agrochemical leaching studies.

The work presented here is part of a larger, ongoing effort, which has a theoretical component and an application for one of the most intensively farmed areas in the United States (the San Joaquin Valley, California). In the theoretical component of the larger study, the optimal set of TTFs is developed and tested with stochastic simulated data for each of the agriculturally relevant soil textures: Sandy Loam, Loam, Silt Loam, Sandy Clay Loam, Clay Loam and Silty Clay Loam. The optimal set of TTFs is subsequently applied, using soils, climate, landuse/pesticide application, and irrigation information to make the first regional-scale, quantitative groundwater vulnerability assessments for the San Joaquin Valley.

An important aspect of TTF model development is to establish the relationship between the number of TTFs in the set, where a reduction of the total number is equivalent to upscaling the TTFs, and the accuracy of the concentration estimates, given the constraints of computational cost. For this reason, the objective of the work reported here was to quantify the effect that upscaling imparts on TTF model concentration estimates in terms of the growth of the errors. For the effort described in this paper, we isolated and evaluated the uncertainty in concentration estimates that relates exclusively to upscaling using the Loam textural class as an example. The results of this study facilitated the selection of the optimal number of TTFs for each soil textural class.

2. Methods

The development of TTF for this study was accomplished with the aid of simulated stochastic data sets for each texture that take the place of observations, as regional-scale pesticide leaching measurements in space and time are difficult, if not impossible, to obtain. The simulated stochastic data sets were built from generated fields of nine distributed soil hydraulic parameters for each soil texture that served as the input to process-based, unsaturated zone fluid flow and conservative solute transport simulations. For example, Figure 1 shows the distribution of the negative log of saturated hydraulic conductivity ($p(K_{sat})$) and porosity (θ_{sat}) for the Loam textural class. For each textural class in the overall study, several sets of TTF with an increasingly greater level of upscaling, as well as one set of TTF for all classes were developed. The development of TTFs involves the estimation of equivalent TF parameters corresponding to a suite of simulated normalized concentration breakthrough curves for a non-reactive tracer. For the study reported here, four sets of TTFs, corresponding to four levels of upscaling were selected for the Loam soil textural class. For each set, the individual TTFs were associated with ranges of porosity and saturated hydraulic conductivity. Solute concentration estimates at the depth of interest for each set of upscaled TTFs were compared to the simulated concentrations from which they were developed according to three sampling schemes. This strategy served to isolate the uncertainty in the TTF estimate resulting from the effects of upscaling only, without introducing uncertainty

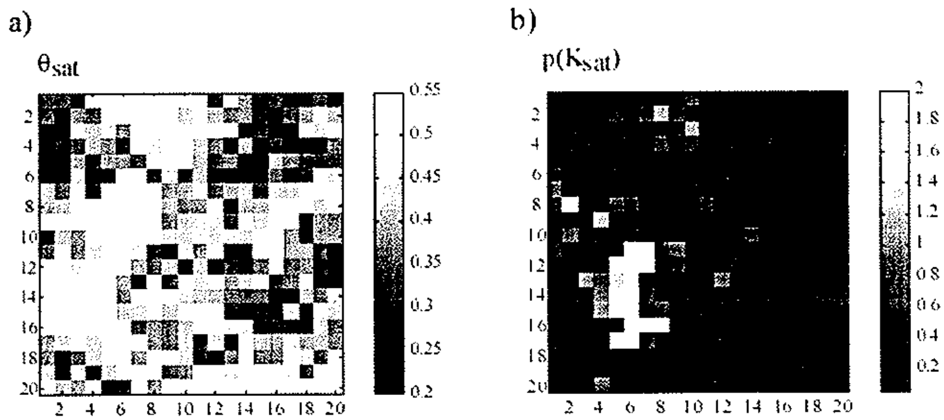
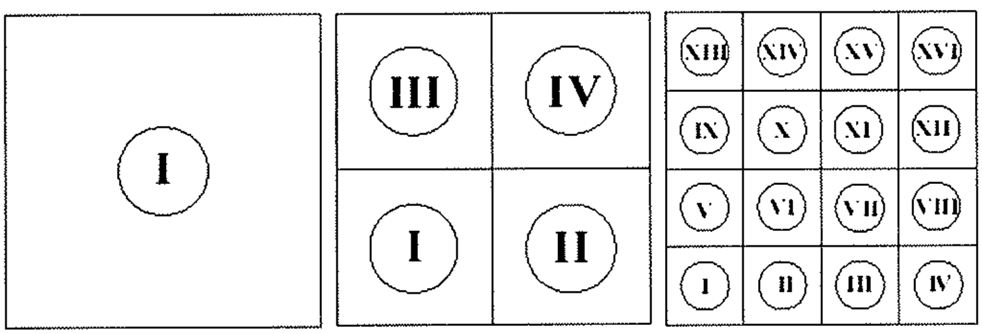


Figure1. Two examples of the generated fields of nine distributed soil hydraulic input parameters for the Loam soil textural class. Each square field has a side length of 200 m, each square element has a side length of 10 m. a) Distribution of porosity (-). b) Distribution of the negative logarithm of saturated hydraulic conductivity (m/day).



Sampling Strategy I   Sampling Strategy II   Sampling Strategy III

Figure 2. The sampling strategies used to seperate the fields shown in Figure 1 into blocks. For each sampling strategy and each set of TTFs, the TTF concentration estimates were compared with the SWMS_3D simulations as a sum total for a given block.

from other sources. Uncertainty was quantified through the sum of the squared error between the concentration time series at a depth of interest as given by the TTF model approach and the process-based simulated stochastic data. The procedure outlined in this section is described in greater detail in the following sections:

2.1 Development of the Stochastic Simulated Data Set for each Soil Texture

The intent of the simulated stochastic data sets is to generate separate distributed parameter fields for each soil textural class that contain all variability in soil hydraulic parameters and therefore leaching behaviour, that is expected to exist within that class. The distributions of the nine most important transport parameters for the each of the six soil textural classes were generated using the sequential Gaussian simulation algorithm contained within the geostatistical software package GSLIB (*Deutsch and Journel, 1998*) and information on spatial structure from our San Joaquin Valley database. For each soil texture, the simulated area was a square area with a side length 200 m and a grid spacing of 10 m. The resulting normal parameter distributions were transformed according to known distributions and their associated means and variances (e.g., *Carsel and Parrish, 1988*) (Figure 1). For each of the six 40,000 m^2 soil textural fields, fully coupled 1-D unsaturated fluid flow and contaminant transport were simulated using the finite-element code SWMS_3D (*Simunek et al., 1995*). Each of the simulations covered a period of one to five years and considered pesticide loading to a compliance surface at 3 m depth. The compliance surface could also be regarded as a surrogate for the water table. The solute concentration time series for each SWMS_3D simulation were used to construct an integrated, regional-scale mosaic of contaminant loading and its spatial and temporal variability at the compliance surface. This regional-scale mosaic represents the 'observations' from which the sets of TTFs are developed and against which they are compared.

2.2. Development of TTF Sets

One breakthrough curve (concentration versus time, measured at the compliance surface) is associated with each of the 400 10 m by 10 m elements that comprise a field with distributed soil hydraulic parameters. Each breakthrough curve can be used to fit the parameters of the Jury CLT described by:

$$C(L,t) = \int_0^t C(0,\hat{o}_{in})\, f(L, t - \hat{o}_{in})\, d\hat{o}_{in}, \tag{1}$$

with:

$$f(L, t - \hat{o}_{in}) = \frac{1}{\sqrt{2\breve{o}}\, \acute{o}\, (t - \hat{o}_{in})} \exp\left[- \frac{(\ln(t - \hat{o}_{in}) - i)^2}{2\acute{o}^2} \right], \tag{2}$$

where C(L, t) is the area-averaged solute flux concentration at depth L and time t, and f(L, t - τ_{in}) is the travel-time probability density function, describing the distribution of solute travel times conditional on the input time τ_{in} (*Jury et al. 1986; Jury and Roth, 1990*) for steady state-conditions. It is assumed that the TF formulation describes the time-invariant leaching behaviour for the site from which it was developed. Actual concentration estimates for a given solute input at this site can be obtained by scaling the TF by the appropriate input concentration, as well as sorption and decay coefficients. Suites of compliance surface breakthrough curves were used to develop several sets of upscaled TTFs. The suites for each set were determined by ranking and grouping what is here termed the TTF-index according to several schemes. The TTF-index is defined as:

$$\text{TTF - index} = \frac{C_p}{(t_p \acute{o})}, \tag{3}$$

where C_p (ML^{-3}) is the concentration peak, t_p (T) is the time at which the peak concentration reaches the depth of interest, and σ (-) is the spread of the travel time distribution.

For this study, the ranked TF-indices are divided into even groups of one, two, four, and eight suites for a given texture. The TTF for each suite is then defined by the upscaled probability density function f_E, with effective parameters μ_E and σ_E, and describes the mean leaching behaviour for the suite:

$$f_E\,(L, t-\tau_{in}) = \frac{1}{\sqrt{2\pi}\,\sigma_E\,(t-\tau_{in})}\exp\left[-\frac{(\ln(t-\tau_{in})-\mu_E)^2}{2\sigma_E^{\,2}}\right]. \qquad (4)$$

Thus, one or more upscaled TTFs for each soil textural class provide a representative leaching response for the range of soil-hydraulic parameters that are typical for that class. Each TTF for each set is associated with a range of saturated hydraulic conductivity (K_{sat}) and saturated soil moisture (θ_{sat}). For the Loam textural class, this procedure yields the sets given in Table 1:

Set	*Number of TTFs in Set*	μ_E	σ_E	K_{sat} *(m/day)*	θ_{sat} (–)
I	1	4.67	0.480	all	all
II	2	4.83	0.500	> 0.10	> 0.42
		4.50	0.488	□ 0.10	□ 0.42
III	4	4.91	0.500	> 0.50	> 0.49
		4.75	0.498	0.10 - 0.50	0.42 - 0.49
		4.61	0.495	0.05 - 0.10	0.36 - 0.42
		4.40	0.487	< 0.05	< 0.36
IV	8	4.97	0.500	> 0.90	> 0.51
		4.86	0.499	0.70 - 0.90	0.47 – 0.51
		4.79	0.496	0.50 - 0.70	0.45 – 0.47
		4.71	0.497	0.30 – 0.50	0.43 – 0.45
		4.65	0.496	0.20 – 0.30	0.38 – 0.43
		4.58	0.494	0.10 – 0.20	0.35 – 0.38
		4.49	0.493	0.05 – 0.10	0.31 – 0.35
		4.30	0.487	< 0.05	< 0.31

Table 1, The sets of TTFs for the Loam textural class considered in this study. Given here are the number of TTFs in each set, the effective parameters μ_E and σ_E for each TTF and the associated ranges of saturated hydraulic conductivity (K_{sat}) and porosity (θ_{sat}).

2.3. Uncertainty Analysis for each set of TTFs

The total uncertainty in simulation estimates results from the composite effect of modeling and parameter uncertainty (*Frey, 1997*). For this study our objective was to isolate the contribution to uncertainty that results from upscaling the TTF model. Therefore, parameters and processes not included in the initial TTF development are not introduced at this stage. The intend of this step was to sample increasingly smaller blocks from the Loam soil textural field and compare the results from the process-based SWMS_3D simulations (the 'observed' values) with the estimates given by each upscaled set of TTFs. The three sampling strategies employed in this study were (Figure 2):

I The whole field was used as one block
II The field was divided into four square blocks of 10x10 elements each
III The field was separated into 16 square blocks of 5x5 elements each

In each block the distribution of K_{sat} and θ_{sat} and the 'true observed' values were assumed to be known. Each set of TTFs was then used to estimate the concentration time series at the compliance surface for each block. From the TTF concentration estimates, the residual error and the sum of the square error (SSQE), as compared to the 'observed' concentration time series, was calculated. Example results of this step are shown in Figure 3.

3. Results

Figure 3a shows the residual error versus time for sampling strategy I, where the concentration at the depth of interest for the entire field was estimated with each of the four sets of TTFs (1, 2, 4, and 8 TTFs in the set). It is clear from the residual time series that (i) the error is time-dependent, with the largest errors occurring in the vicinity of the concentration peaks, and (ii) the TTF predictions are unbiased with no systematic over- or underprediction of concentrations. As one would expect, the SSQE decreases with the number of TTFs in the set for sampling strategy I (Figure 3b). However, it is important to note that the largest improvement (by a factor of 7) in the SSQE is achieved when increasing the number of TTFs in the set from one to two. Larger sets of TTFs result in only small subsequent improvement of TTF concentration estimates when compared to the first set that contains only one TTF.
The impact of using upscaled TTFs is less clear, when increasingly smaller areas of the field are considered for sampling strategy II (Figure 3c) or sampling strategy III (Figure 3d). The subsets of elements contained in each block for sampling strategies II and III are less likely to span the variation of soil hydraulic parameters that is encountered in the entire field. Therefore it is reasonable to anticipate that any set of TTFs is increasingly capable of describing the average leaching behaviour of the block, resulting in a decreased SSQE for all sets and blocks. In fact, the SSQE is one and two orders of magnitude smaller for sampling strategies II and III respectively (Figure 3c,d). Generally, as expected, the SSQE decreases with the number of TTFs in the set. As before, the most dramatic decrease takes place when increasing the number of TTFs in the set to two. However, for several blocks a non-intuitive increase of the SSQE with the number of TTFs in the set is observed (i.e., Figure 3c, Block I). This increase is due to the fact that in this study TTF selection within a given set is tied only to θ_{sat} and K_{sat}. However, there are a total of nine soil hydraulic parameters that serve as input for the process-based SWMS_3D simulations for each element. All nine parameters interdepently influence the leaching behaviour for a given element. Since the remaining seven parameters do not exert any influence on TTF selection within a set, a seemingly contradictory increase of the SSQE with the number of TTFs results in some cases. Considering the size of the errors, the increased complexity of the TTF approach, and the additional data requirements that incorporating information about the other seven parameters into TTF selection would entail, we believe that the current protocol represents the best compromise between the accuracy of concentration estimates and model simplicity.

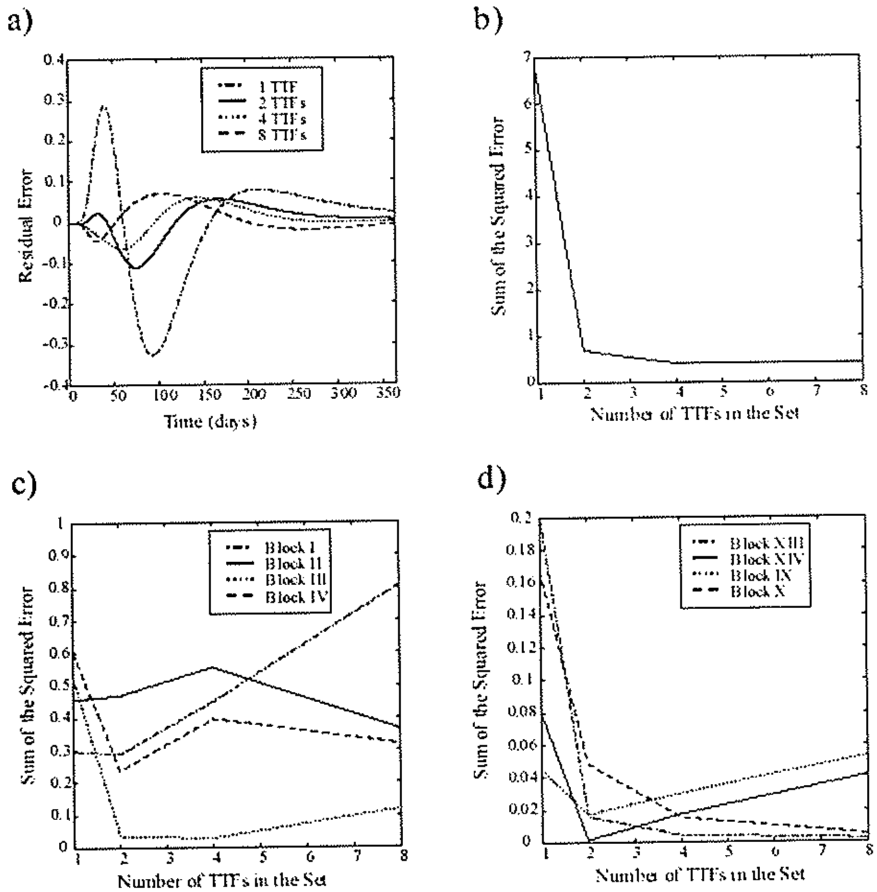


Figure 3. Results from the comparison of the TTFs solute concentration estimates with the process-based SWMS_3D simulations for all four sets of TTFs and the three sampling strategies. a) Sampling strategy I (1 block): The resisual error (g/m^3) for all four sets of TTFs, b) Sampling strategy I (1 block): The sum of the squared error $(g/m^3)^2$ versus the set of TTFs, c) Sampling strategy II (4 blocks): The sum of the squared error $(g/m^3)^2$ versus the set of TTFs for all four blocks, d) Sampling strategy III (16 blocks): The sum of the squared error $(g/m^3)^2$ versus the set of TTFs for the four blocks in the upper left quadrant (equal to Block III of sampling strategy II). Please note the different vertical scales in b), c), and d).

4. Conclusions

This study investigated the uncertainty in solute concentration estimates that is introduced by upscaling TTFs with the use of three sampling strategies and four sets of TTFs that represent increasing levels of upscaling. Uncertainty was quantified in terms of the growth of the errors. It is clear from our results that (i) all sets of TTFs are capable of representing leaching behaviour for all sampling strategies of this study with reasonable accuracy, (ii) TTF predictions of solute concentrations improve for areas with less variability in soil hydraulic parameters (smaller block size in this study), (iii) uncertainty in TTF estimates is significantly reduced when using the set containing two TTFs as compared to the upscaled set containing one. The uncertainty in TTF estimates was not consistently reduced further by adding more TTFs to the set.

Acknowledgements

The work described here is a CESIR (Center for Earth Science Information Research) contribution. The first author gratefully acknowledges the support for this work from an EPA STAR fellowship and a Theresa Heinz Fellowship for Environmental Research. The computing facilities used in this study were provided by an National Science Foundation equipment grant to Steve Gorelick and Keith Loague.

References

- *Carsel, R.F., Parrish, R.S., 1988,* Developing joint probability distributions of soil-water retention characteristics, Water Resources Research 24, pp. 755-769.
- *Corwin, D.L., Loague, K., Ellsworth, T.R., 1999*, Introduction: Assessing non-point source pollution in the vadose zone with advanced information technologies, in: Assessment of Non-Point Source Pollution in the Vadose Zone, Corwin, D.L., Loague, K., Ellsworth, T.R., Eds., AGU Geophysical Monograph 108, pp. 1-20.
- *Deutsch, C.V., Journel, A.G., 1998,* Geostatistical Software Library and User's Guide. Oxford University Press, New York.
- *Frey, H.C., 1997,* Quantitative Analysis of Variability and Uncertainty in Energy and Environmental Systems, in: Uncertainty Modeling and Analysis in Civil Engineering, Ayyub, B.M., Ed., CRC Press, pp. 381-423.
- *Jury, W.A., 1982*, Simulation of solute transport using a transfer function model, Water Resources Research 18, pp. 363-368.
- *Jury, W.A., Stolzy, L.H., Shouse, P., 1982*, A field test of the transfer function model for predicting solute transport, Water Resources Research 18, pp.369-375.
- *Jury, W.A., Roth, K., 1990*, Transfer functions and solute movement through soil. Birkhaeuser Verlag, Basel, Switzerland.
- *Simunek, K., Huang, K., van Genuchten, M., Th., 1995*, The SWMS_3D code for simulating water flow and solute transport in three-dimensional variably-saturated media. USSL/USDA Report No. 139. Riverside, CA.
- *Stewart, I.T., Loague, K., 1999*, Identification of "type" transfer functions for regional-scale assessment of non-point source groundwater contamination in Fresno County, California. Journal of Environmental Quality 28, 378-387.
- *White, R.E., Dyson, J.S., Haigh, R.H., Jury, W.A., Sposito, G., 1986*, A transfer function model of solute transport through soil, 2. Illustrative applications, Water Resources Research 22, pp. 248-254.