XVIII International Conference on Water Resources
CMWR 2010
J. Carrera (Ed)
CIMNE, Barcelona 2010

DECISION SUPPORT BASED ON UNCERTAINTY QUANTIFICATION OF MODEL PREDICTIONS OF CONTAMINANT TRANSPORT

Velimir V. Vesselinov* **and Dylan R. Harp***

*Computational Earth Science Group, Earth and Environmental Sciences Division
Los Alamos National Laboratory, MS T003, Los Alamos NM 87545, U.S.A.
e-mail: vvv@lanl.edu, web page: http://www.ees.lanl.gov/staff/monty

Key words: decision support, uncertainty analysis, inverse analysis, model predictions, source identification, monitoring network design, optimization, constrained calibration

Summary. The process of decision making to protect groundwater resources requires a detailed estimation of uncertainties in model predictions. Various uncertainties associated with model development, such as measurement and computational errors, uncertainties in the conceptual model and model-parameter estimates, simplifications in model setup and numerical representation of governing processes, influence the uncertainties in the model predictions. As a result, the predictive uncertainties are generally difficult to quantify. Quite frequently however, the uncertainties in only some of the model parameters and predictions are important to consider in the decision making process. We investigate and compare existing and newly-proposed methods for the quantification of predictive uncertainties in relation to decision support. The goal is to quantify predictive uncertainties affecting decision making related to locating new monitoring wells.

1 INTRODUCTION

Typically, a process of decision making related to protection of groundwater resources requires detailed estimation of uncertainties in model predictions. Various uncertainties associated with model development, such as measurement and computational errors, uncertainties in conceptual model and model-parameter estimates, simplifications in model setup and numerical representation of governing processes, influence the uncertainties in the model predictions. As a result, the predictive uncertainties are generally difficult to quantify. Here we present analyses of model uncertainties in a decision support framework by evaluating the addition of supplementary monitoring wells to an existing monitoring network. The monitoring wells needs to be located in a way to reduce uncertainty in model predictions of (1) contaminant mass and its current and future spatial distribution in the aquifer and (2) location of the unknown contaminant source.

An important aspect of the presented analysis is the identification of a contaminant source within an aquifer based on contaminant-concentration observation data. Source identification includes estimation of various spatial and temporal characteristics such as location, size, and transients in contaminant flux. Various source-identification approaches have been presented in the literature[1,2,3]. Here we apply a novel probabilistic approach that couples inverse

analyses (model calibration) with uncertainty analysis of model predictions. The model simulation of contaminant fate and transport is performed forward in time from multiple potential entry points at the regional aquifer. Some of the benefits of using this approach compared to other existing methods ([3] provides detailed summary) are related to improved numerical stability and computational efficiency. The approach is applied to characterize an existing contaminant (chromium) plume at the regional aquifer beneath Sandia Canyon, Los Alamos National Laboratory.

2 METHODOLOGY

The analyses require a model capable of simulating three-dimensional contaminant transport in the regional aquifer. An analytical model is applied to simulate contaminant transport in the regional aquifer. The simulations do not include groundwater transport in the vadose zone above the regional aquifer; however, uncertainties in the vadose-zone transport are implicitly represented by uncertainty in the contaminant arrival location and mass flux. The effects of geochemical interactions of contaminants with rocks along the flow paths are also not considered. The analytical solution for concentration $c(x,y,z,t)$ of a contaminant released at a source with a rectangular-parallelepiped shape in three-dimensional uniform semi-bounded aquifer has the following form[4,5,6]:

$$c(x,y,z,t) = \frac{1}{8\pi\theta x_S y_S z_S}\int_0^t I(t-\tau)\exp(-\lambda\tau)\left[erfc\left(\frac{x-\frac{1}{2}x_S-v\tau}{2\sqrt{\alpha_L v\tau}}\right)-erfc\left(\frac{x+\frac{1}{2}x_S-v\tau}{2\sqrt{\alpha_L v\tau}}\right)\right]\times$$
$$\left[erfc\left(\frac{y-\frac{1}{2}y_S-v\tau}{2\sqrt{\alpha_{TH} v\tau}}\right)-erfc\left(\frac{y+\frac{1}{2}y_S-v\tau}{2\sqrt{\alpha_{TH} v\tau}}\right)\right]\times\left[erfc\left(\frac{z-(z_0+z_S)}{2\sqrt{\alpha_{TV} v\tau}}\right)-erfc\left(\frac{z-z_0}{2\sqrt{\alpha_{TV} v\tau}}\right)+\right. \quad (1)$$
$$\left.erfc\left(\frac{z+(z_0+z_S)}{2\sqrt{\alpha_{TV} v\tau}}\right)-erfc\left(\frac{z+z_0}{2\sqrt{\alpha_{TV} v\tau}}\right)\right]d\tau$$

where: α_L, α_{TH}, α_{TV} are the longitudinal, transverse horizontal and transverse vertical dispersivities [L], respectively; λ is the half-life decay constant [T^{-1}], θ is effective transport porosity [-], and $I(t)$ is the transient contaminant mass flux [MT^{-1}]. The groundwater flow occurs along the *x* axis with Darcy velocity *v* [*L/T*]. The source dimensions are x_S, y_S and z_S [L] along each axes, respectively. The source center is located at $(0, 0, z_0 + \frac{z_S}{2})$; if $z_0 = 0$, the source location will be at the top of the aquifer. The aquifer is bounded at the top ($z = 0$) and has infinite thickness. The numerical solution of (1) utilizes functions provided in the GNU Scientific Library (GSL)[7]. Various codes for simulation and decision-making related to contaminant transport (e.g. Art3d, Bioscreen) currently use an analytical solution[8] that has been shown to be mathematically inaccurate, especially when contaminant transport is dispersion dominated[9,10]; Eq. 1 has not been found to suffer from this issue.

In the analyses presented below, the contaminant mass flux is defined as:

$$I(t) = \begin{cases} if\ t < t_0 & 0 \\ if\ t > t_0\ AND\ t < t_1 & f \\ if\ t > t_1 & 0 \end{cases} \quad (2)$$

where a steady contaminant mass flux f [MT^{-1}] is introduced within a specified time window (t_0, t_1). To account for uncertainty in the advective-transport flow direction, Eq. 1 is solved for $c(x', y', z', t)$ and the coordinate system is rotated laterally using the following expression to compute $c(x, y, z, t)$:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} cos\alpha & -sin\alpha & 0 \\ sin\alpha & cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} \quad (3)$$

where α defines the angle of rotation.

Compared to numerical models, the applied analytical solution for the simulation of three-dimensional contaminant transport has limitations related to some of its assumptions: (1) the aquifer has uniform properties; and (2) groundwater flow is steady-state and uniform. However, it allows for (1) fast simulation times (less than a second), (2) no numerical or resolution accuracy issues, and (3) simple incorporation of transients in the input contaminant flux and the observed concentrations. It is important to note that the model runtime is critical for the performance of the analyses presented below; the estimation of the potential locations of contaminant arrivals requires on the order of 10^5 to 10^6 analytical model executions. A single execution of a numerical model at the same scale and with similar complexity (uniform properties and groundwater flow) requires 10 to 60 minutes.

To perform uncertainty analysis of model predictions and estimate potential locations of contaminant arrival, the analytical model is calibrated against existing data measurements. The calibration is performed using various optimization techniques (global and gradient-based). Various objective functions have been explored in the optimization process. The analyses demonstrated that it is critical to account for the uncertainty in the calibration targets. The following objective function appeared to produce the most promising results:

$$\Phi(\boldsymbol{p}) = \sum_{i=1}^{M} \begin{cases} if\ c_i(\boldsymbol{p}) > \bar{c}_{i,max} & \left[\bar{c}_{i,max} - c_i(\boldsymbol{p})\right]^2 \\ if\ c_i(\boldsymbol{p}) < \bar{c}_{i,max}\ AND\ c_i(\boldsymbol{p}) > \bar{c}_{i,min} & 0 \\ if\ c_i(\boldsymbol{p}) < \bar{c}_{i,min} & \left[\bar{c}_{i,min} - c_i(\boldsymbol{p})\right]^2 \end{cases} + \sum_{i=1}^{M} \left|\left(\bar{c}_i - c_i(\boldsymbol{p})\right)\right| \quad (4)$$

where $\boldsymbol{p}$ is a vector of calibrated model parameters (N elements), M is the number of calibration targets in vector $\bar{\boldsymbol{c}}$, $\bar{\boldsymbol{c}}_{min}$ and $\bar{\boldsymbol{c}}_{max}$ define an acceptable ranges for model predicted concentrations $\boldsymbol{c}(\boldsymbol{p})$. The objective function has two terms: the first quadratic term accounts for deviations of model prediction from the acceptable ranges; the second linear term accounts for model deviations from the best estimate of the calibration targets.

Due to various uncertainties, the optimization problem associated with Eq. 4 is ill-posed.

Model parameter	Prior estimate	Prior uncertainty		Posterior estimate	Posterior uncertainty	
		Minimum	Maximum		Minimum	Maximum
Source x coordinate [m]	499174	498400	499150	499150	498450	499350
Source y coordinate [m]	539122	538700	539550	539000	538850	539550
Source longitudinal dimension [m]	100	10	500	27	10	498
Source transverse dimension [m]	100	10	500	468	10	500
Contaminant flux f [kg/a]	10	1	1000	18	10	78
Contaminant arrival time t_0 [a]	1990	1956	2004	1977	1960	1994
Flow Angle θ [°]	-35	-75	0	-7	-75	0
Pore velocity [m/a]	50	1	600	23	3	106
Longitudinal dispersivity [m]	100	1	200	1	1	99
Horizontal transverse dispersivity [m]	10	1	100	6	1	100
Vertical transverse dispersivity [m]	1	0.1	10	1	0.10	5
Total mass [kg] (computed)	-	-	-	550	270	3300

Table 1: Model parameters: initial estimates and calibration results

Due to (1) substantial prior uncertainties, (2) large number of unknown parameter, (3) limited amount of observation data and (4) model complexity and non-linearity, multiple plausible solutions of the inverse problem appear to be feasible. To explore the existing uncertainties in model predictions various alternative approaches have been tested. These included global optimization (e.g. particle swarm) and Monte Carlo methods (including Null-Space and Markov-Chain techniques). The most computationally efficient method was based on inverse analyses utilizing partial parameter-space discretization. In this case, prior uncertainty ranges for some of the model parameters are uniformly discretized. Multiple inversions are performed for each combination of discretized parameters keeping these parameters fixed at their discrete values during optimization process. In the analyses presented below, the x and y coordinates of the source location are discretized defining a set of plausible contaminant arrival locations. The simulation and optimization modules are part of a program package called MODUL[12].

The goals are to (1) calibrate the analytical model of contaminant transport against observed concentration data, (2) estimate potential locations of contaminant arrival based on all the available site data, (3) estimate spatial and temporal (past and current) distributions of contaminant mass in the regional aquifer and their associated uncertainties. This information will be applied for model-based decision-making for placing an additional monitoring well.

3 MODEL SETUP

The parameters applied in model simulations are listed in Table 1. The initial (prior) estimates and prior uncertainty ranges are estimated based on the available site-specific and published data[11]. The prior uncertainty ranges are defined to be within a wide range to account for the existing uncertainty (Table 1). The properties of contaminant arrival location at the top of the regional aquifer are uncertain. Contaminants are expected to arrive at an unknown lateral location at the top of the regional aquifer (z_0=0) and within an unknown lateral dimension but fixed vertical size (z_S=1 m). The source thickness represents a zone of

initial mixing of contaminants arriving from the vadose zone in the regional aquifer groundwater; the zone is also impacted by seasonal fluctuations of the regional water table (~0.2 m). Contaminant mass flux is assumed steady and equal to unknown constant value within a specified time window (t_0, t_1). The time of arrival of contaminants at the top of the aquifer, t_0, is estimated to vary between 1956 and 2004 to accommodate the existing uncertainty. The time at which contaminants cease to arrive in the regional aquifer, t_1, is fixed (2020) and does not influence the simulations performed below; the simulations represent present day concentrations (t=2009). The contaminant half-life decay constant (λ) is set to zero since existing data suggest no geochemical reduction of contaminants (Cr^{6+}) occurs in the regional aquifer.

The initial estimate (50 m/a) and uncertainty range (1-600 m/a) of the pore velocity (Table 1) are based on available site-specific data about permeability and hydraulic gradients[11]. There is uncertainty in the hydraulic gradient due to data scarcity and measurement uncertainties; the best estimate is that the horizontal gradient is at an angle of around -30° (measured from the x axis). The sediments in the area are highly stratified including sequences of nearly horizontal layers with contrasting hydraulic properties; site-scale aquifer permeability is expected to exhibit anisotropy due to the structure of the heterogeneity[11]. Horizontal and vertical anisotropy (the permeability tensor) of the medium are poorly characterized. Field data suggest vertical permeability orders of magnitude lower than the lateral permeability[11]. The dominant principal direction of the permeability tensor is expected to be horizontal and oriented at an angle of about -45°. To account for existing uncertainties, the contaminant flow directions are allowed to vary widely to cover a broad range of advective-flow direction (from 0° to -75°; Table 1). The impact of vertical anisotropy on contaminant transport is captured through selection of the dispersivity coefficients (transverse vertical dispersivity vs. longitudinal and transverse horizontal dispersivities). There are no site-specific estimates of the dispersivity coefficients; initial estimates and uncertainty ranges are based on published data[11] that take into account the scale at which groundwater transport occurs.

The contaminant concentrations in the regional aquifer observed at the monitoring wells are the calibration targets in the inversion. These contaminant concentrations are uncertain due to: (1) measurement errors, (2) uncertainty in background contaminant concentration (5-7 ppb; background concentrations are subtracted from the observed concentrations to obtain calibration targets; the model simulates concentrations above the background concentrations), and (3) seasonal fluctuations in contaminant concentration which might be caused by processes that are currently not represented in the simulation process. The calibration data are presented in Table 2. For all the wells except R-28, the calibration targets are equal to average concentrations representing current data (circa 2009); two calibration targets (circa 2004 and 2009) are applied for R-28. The objective function (Eq. 4) is computed based on concentrations listed in Table 2.

Well #screen	x [m]	y [m]	Screen z [m]		t [a]	Contaminant concentration (Cr^{6+}) [ppb]					
						Observed			Model-predicted		
			top	bot-tom		Best	Uncertainty		Best	Uncertainty	
							Min	Max		Min	Max
R-11	499882.6	539296.1	5.6	12.6	2009	13	0	30	13	0	27
R-13	500174.4	538579.8	36.7	55.1	2009	5	0	15	0	0	13
R-15	498442.1	538969.5	0.0	15.0	2009	10	0	30	0	0	22
R-28	499563.7	538995.8	13.2	20.4	2004	400	300	700	400	320	400
					2009	400	300	700	400	390	470
R-34	500972.0	537650.7	26.7	33.7	2009	5	0	15	0	0	0
R-35a	500581.1	539286.0	69.0	84.0	2009	5	0	15	0	0	0
R-35b	500553.2	539289.6	11.2	18.2	2009	8	0	15	0	0	0
R-36	501062.9	538806.1	4.9	11.9	2009	10	0	15	0	0	0
R-42	499174.0	539122.8	3.7	10.1	2009	850	600	1000	850	790	880
R-43#1	499029.6	539378.6	3.3	9.6	2009	4	0	10	0	0	9
R-43#2	499029.6	539378.6	23.2	26.2	2009	3	0	10	0	0	5
R-44#1	499890.7	538615.1	4.9	8.0	2009	12	0	15	9.9	0.1	14
R-44#2	499890.7	538615.1	32.5	35.5	2009	5	0	10	6	0.1	10
R-45#1	499948.1	538891.8	3.6	6.6	2009	12	0	15	11	0	16
R-45#2	499948.1	538891.8	32.5	38.6	2009	5	0	10	7	0	11
R-50a	499461.3	538599.7	6	12	2009	-	-	-	0	0	290
R-50b	499461.3	538599.7	6	12	2009	-	-	-	0	0	50
R-50c	499461.3	538599.7	6	12	2009	-	-	-	0	0	12

Table 2: Calibration targets and modeling results

3 RESULTS AND DISCUSSION

Inverse analyses are performed for each potential contaminant-arrival location presented in Fig. 1. If the model is successfully calibrated, i.e. model-predicted concentrations are within the calibration ranges (Table 2), the potential contaminant-arrival location is defined to be plausible. The identified plausible locations of contaminant arrival at the top of the regional aquifer are shown in Fig. 2. The spatial discontinuity of the identified locations demonstrates complexity of the explored parameter space. Out of 551 potential locations, the analysis has identified 54 as plausible. It is important to note that previous analyses produce similar results[11] though they were based on a smaller calibration data set and using different methodology and computational approach. This demonstrates consistency in the analyses, and a general trend of diminishing uncertainty in the model predictions with the collection of new data. Fig. 2 presents a representative subset (11 out of 54) from all the plausible model-predicted contaminant plumes. Plumes are plotted showing concentrations higher than 50 ppb. The plumes are consistent with all the available hydrogeological and geochemical data. They are produced using models that predict contaminant transport from one of the contaminant-arrival locations presented in Fig. 1. Each model has parameters that are consistent with

accepted uncertainty ranges of model parameters (Table 1) and uncertainty ranges of contaminant concentrations at the monitoring wells (Table 2). Fig. 3 shows the best estimate of the plausible contaminant concentrations along the regional water table circa 2009. The best estimate is computed by averaging the concentrations of all the of plausible model-predicted contaminant plumes.

Table 1 lists the best estimates and estimated uncertainty ranges associated with the successfully calibrated analytical models. It is important to note that the calibration process did not provide a substantial reduction in the uncertainty associated with model parameters. Table 2 lists the observed and model predicted concentrations (best estimates and uncertainty ranges) for existing monitoring wells. Based on the model analyses, the contaminant mass that currently resides in the regional aquifer can be estimated. Accounting for existing uncertainties in the model parameters and calibration targets, the contaminant mass in the regional aquifer is estimated to be between 270 and 3300 kg (best estimate ~550 kg; Table 1).

The optimization process has been applied to estimate the reduction in the uncertainty of model predictions of the spatial extent of the plume if a new well is added to the existing monitoring network. Fig. 3 shows that best estimate for the new well location (labeled R-50). Based on the performed analyses, the contaminant concentrations at the new well are expected to be at background levels; however, there is substantial uncertainty in the model predictions (concentration can be as high 300 ppb; Table 2).

4 CONCLUSIONS

We have applied various approaches to explore the existing uncertainties in model predictions for the scenario presented here. Levenberg-Marquardt optimization utilizing partial parameter-space discretization appears to provide the best performance in terms of both analytical and computationally efficient. The calibration process did not provide substantial reduction in the uncertainty in the model parameters. Nevertheless, the analyses effectively constrained the plausible locations of contaminant arrival as well as the model predicted spatial distribution of contaminant mass in the aquifer. The results provided support for decision making related to locating a new monitoring well.

REFERENCES

[1] J. Atmadja, and A.C. Bagtzoglou, State-of-the-art report on mathematical methods for groundwater pollution source identification, *Environ. Forensics*, **2**(3): 205-214 (2001).

[2] R.M. Neupauer, R. Lin, and H. O'Shea, Conditioned backward probability modeling to identify sources of groundwater contaminants subject to sorption and decay, *Water Resour. Res.*, **43**, W11403, doi:10.1029/2006WR005580 (2007).

[3] Z.Dokou, and G.F. Pinder, Optimal search strategy for the definition of a DNAPL source, *Journal of Hydrology*, **376**, 542-556 (2009).

[4] E.J. Wexler, *Analytical solutions for one-, two-, and three-dimensional solute transport in ground-water systems with uniform flow*, U.S. Geological Survey, Open-file Rep. 89-56 (1992)

[5] E. Park, H. Zhan, Analytical solutions of contaminant transport from finite one-, two-,

and three-dimensional sources in a finite-thickness aquifer, *Journal of Contaminant Hydrology*, **53**, 41–61 (2001).

[6] H. Wang and H. Wu, Analytical solutions of three-dimensional contaminant transport in uniform flow field in porous media: A library, *Journal of Frontiers of Environmental Science & Engineering in China*, **3**, 112-128 (2009)

[7] GSL: GNU Scientific Library, http://www.gnu.org/software/gsl

[8] P.A. Domenico, An analytical model for multidimensional transport of a decaying contaminant species. *Journal of Hydrology* **91**, 49–58 (1987).

[9] M.R. West, B.H. Kueper, and M.J. Ungs. On the Use and Error of Approximation in the Domenico (1987) Solution, *Ground Water*, **45**(2), 126-135 (2007).

[10] V. Srinivasan, T.P. Clement and K.K. Lee, Domenico solution--is it valid? *Ground Water*, **45**(2), 136-46 (2007).

[11] Los Alamos National Laboratory, *Sandia Canyon Investigation Report*, **LA-UR-09-5123,** Los Alamos, New Mexico (2009).

[12] V.V. Vesselinov, MODUL: MOdel-based Decision-support and Uncertainty analysis Library, http://www.ees.lanl.gov/staff/monty/codes/modul (2009).

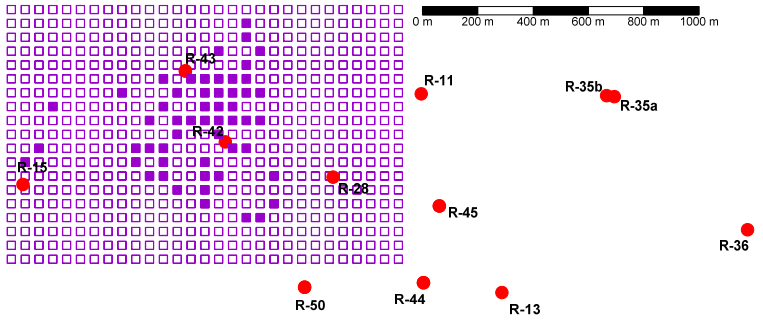

Figure 1: 2D grid (29 x 19) of 551 potential contaminant-arrival locations from the vadose zone to the top of the regional aquifer; the locations are spaced 50 x 50 m (squares do not represent the source sizes). Identified potential source locations (54 out of 551) are shown as solid squares. Rejected source locations are shows as open squares

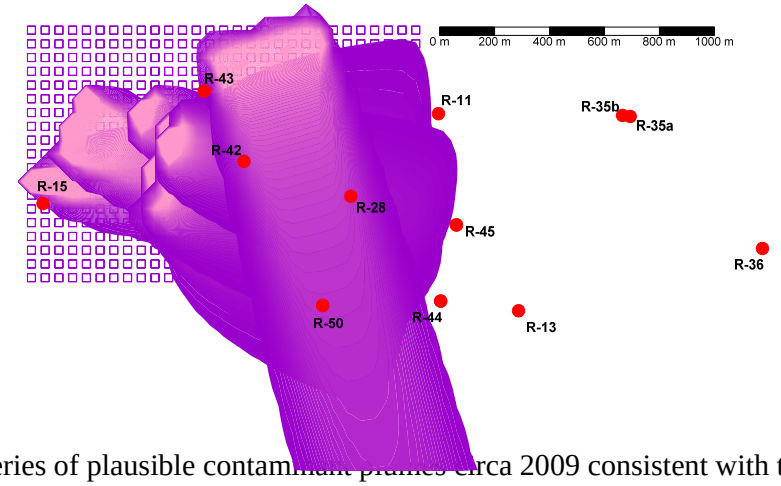

Figure 2: Series of plausible contaminant plumes circa 2009 consistent with the available hydrogeological and geochemical data (plumes are plotted showing c> 50 ppb).

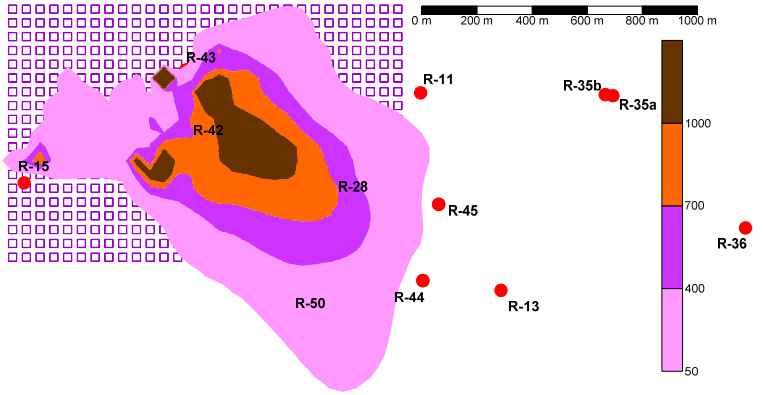

Figure 3: Best estimate of the contaminant concentrations along the regional water table circa 2009.