Hypothesis testing of buoyant plume migration using a highly parameterized variable-density groundwater model at a site in Florida, USA

Alyssa M. Dausman · John Doherty · Christian D. Langevin · Joann Dixon

Abstract A highly parameterized variable-density groundwater flow and solute transport model was developed to test multiple hypotheses for upward movement of treated wastewater (effluent) injected into a saline coastal aquifer in southeastern Florida, USA. The model was designed to assess risk to a drinking-water aquifer above the zone of injection, where monitoring wells have detected effluent. The model-based analysis accommodated geological and data complexity, including the observed presence of effluent in upper monitoring wells, but not in lower monitoring wells, thereby giving the appearance of the effluent having bypassed geological layers. The modeling approach included the application of multiple methodologies to reduce model run times during parameter estimation while providing detailed calibrated model(s) that can be used to assess the potential capacity for different mechanisms of effluent migration. The methods included use of a semi-analytical equation to quickly calculate initial concentrations, parallelization of model runs over multiple processors when calibrating, and utilization of the concepts of singular value decomposition and Tikhonov regularization to accommodate a high level of parameterization complexity. The results reveal that vertical effluent migration could occur as diffuse flow through heterogeneous confining units; however, flow through a channelized pathway caused by well construction appears to be more likely.

Keywords Variable-density · Heterogeneity · Regularization · Parameterization · USA

Introduction

One or more hypotheses are typically proposed to explain observations. In groundwater studies, a "phenomenon" in observation data can be unexpected contamination of a well field or unexplained data collected from groundwater wells such as highly varying water levels or the appearance of an unexplained solute constituent. Numerical models have historically been developed to understand groundwater problems and provide insight into aquifer parameters and system properties. Models can also be used to test hypotheses pertaining to groundwater movement and processes. Remson et al. (1980) undertook model-based hypothesis testing to identify aquifer characteristics and determine groundwater recharge in California, interpret data in Utah, and explore water resources in Taiwan. Pennequin (1983) used a numerical model to determine hydrogeological characteristics and understand the groundwater flow component to a lake in Wisconsin where multiple hypotheses were presented and tested to explain data collected in the area. Multiple geological configurations were also tested with a groundwater model to understand hydraulic conductivity distributions in Canada (Jamieson and Freeze 1983). More recently, Krabbenhoft and Anderson (2006) tested multiple hypotheses with a numerical model to explain groundwater data collected beneath Trout Lake in Wisconsin and understand groundwater seepage through the lakebed.

Carrera et al. (2005) pointed out that multiple conceptual models with different sets of parameters can often replicate the same set of geological and system observations reasonably well; on these occasions more than one hypothesis pertaining to groundwater processes and movement can be retained. Bredehoeft (2005) shows that if the existence of multiple working hypotheses is ignored, model-based

Received: 23 December 2008 / Accepted: 1 August 2009
Published online: 3 October 2009

A. M. Dausman (✉) · C. D. Langevin · J. Dixon
US Geological Survey,
3110 SW 9th Ave., Fort Lauderdale, FL 33315, USA
e-mail: adausman@usgs.gov
Tel.: +954-377-5972
Fax: +954-377-5901

C. D. Langevin
e-mail: langevin@usgs.gov

J. Dixon
e-mail: jdixon@usgs.gov

J. Doherty
Watermark Numerical Computing Pty. Ltd.,
336 Cliveden Ave., Corinda 4075, Australia
e-mail: johndoherty@ozemail.com.au

 DOI 10.1007/s10040-009-0511-6

uncertainty analysis may in fact underestimate the degree of uncertainty associated with certain model predictions. This is particularly true when making long-term predictions. Troldborg et al. (2007) found that multiple (conceptual) models could be calibrated to the same set of water level and discharge data, this being the main source of uncertainty affecting future predictions of system behavior, particularly contaminant transport. Ye et al. (2008) evaluated different statistical methods by which to rank the comparative likelihood of competing conceptual models and the uncertainty related to them. If more than one hypothesis and/or conceptual model are retained after others are rejected because of an inability to adequately fit historical field data, model predictions should be made with all retained conceptual models. In general, different predictions could be made by these multiple retained models at different future times. The predictions can then be used as a basis for additional data collection that may help either refine or eliminate hypotheses and/or conceptual models.

Groundwater model-based hypothesis testing can be implemented through attempting to fit historical data on the basis of different depictions of a modeled area. If a satisfactory fit with historical observations cannot be made on the basis of a given hypothesis, or if the cost of obtaining a satisfactory fit is the introduction of unrealistic characteristics to the calibrated parameters, then the hypothesis can be rejected. Hypotheses that cannot be rejected are retained; retained parameter fields then constitute a basis for predicting future conditions and for interpolation between current observation locations. Where various and very different hypotheses are tested through sequential model re-calibration in this manner, it is important to ensure that any particular hypothesis is not falsely rejected through an inability on the part of the model to simulate system hydrogeological characteristics. Even groundwater systems that are characterized by relatively simple geology can still contain a high degree of natural heterogeneity; recognizing the possible existence of this heterogeneity is fundamental to the simulation of contaminant transport. Therefore, it is important that a model contains a level of parameterization density that is compatible with the potential spatial variability of salient hydrogeological properties in order to prevent artificial suppression of the effects of that variability on contaminant movement (Carrera et al. 2005). With this approach to parameterization, calibration-based hypothesis testing is able to introduce variability into parameter fields that may lead to multiple model configurations that support existing water level and transport data.

Existing applications of groundwater models for hypothesis testing have focused on single-density groundwater systems. In this paper, the concept of hypothesis testing is applied to a complicated variable-density problem that involves the buoyancy-driven upward migration of treated wastewater (effluent) injected into a deep saline aquifer; observations indicate that the effluent is migrating vertically toward a protected drinking water aquifer. A three-dimensional, variable-density groundwater flow and solute transport model was developed to facilitate the hypothesis testing. The groundwater model is highly parameterized to accommodate the spatial variability required to match historical migration patterns. The computationally demanding nature of this problem required efficient strategies for generating internally consistent initial heads and salinities, capturing heterogeneity at different spatial scales, and calibrating multiple highly parameterized models. The calibrated models achieved through the hypothesis testing procedure described here are then used to explore the likelihood of various mechanisms for the unexpected effluent migration.

Site history

The South District Wastewater Treatment Plant (SDWWTP), located in southeastern Florida (Fig. 1), is the fifth largest capacity deep-well injection plant in the world. Installation of injection and monitoring wells at the site began in 1977, and was completed by 1996. The first injection well, IW-5, was completed in October 1977. Construction of well IW-5 was followed by the construction of a further 48 injection and monitoring wells (Fig. 2). Injection and monitoring wells are constructed using a staged casing system that monitors distinct zones (CH2M Hill 1981). Effluent injection began at the site in 1983 at a rate of approximately 1.9 x 10^8 L/day (liters per day). By 2007, the site was permitted to inject approximately 3.8×10^8 L/day. Injection of treated effluent is into the Lower Floridan Aquifer (LFA; Fig. 3), a carbonate aquifer system approximately 700 m below sea level (BSL). With the exception of IW-8 to IW-17, injection wells are cased from the land surface down to the top of the LFA. IW-8 to IW-17 were cased above the Delray Dolomite because of difficulties encountered during drilling (CH2M Hill 1981). Monitoring wells constructed at the site are designed to detect the upward migration of treated effluent and to monitor changes in water levels in the Floridan aquifer system (CH2M Hill 1977). The BZ monitoring well is open to four different zones at depths of 311, 495, 750, and 820 m. The remaining monitoring wells, with the exception of FA-4 and FA-14, are each open to two different zones: an upper zone and a lower zone. For most monitoring wells, the upper zone is in the Middle Floridan Aquifer (MFA); the lower zone is in the Middle Confining Unit 2 (MCU2). FA-4 and FA-14 monitor only one zone.

Ammonia concentrations exceeding ambient conditions have been detected in both lower and upper monitoring zones suggesting effluent has migrated upward at the site (Figs. 2, 3, and 4; Maliva et al. 2007). Ammonia concentrations are used here as an indicator of effluent. The relative effluent fraction is calculated as the measured ammonia concentration divided by the maximum injected ammonia concentration. Effluent plumes have been detected in the northwest (NW), northeast (NE), and southeast (SE) areas of the site. An interesting complexity in the plume characteristics is that the NE and NW plumes have been detected in the upper monitoring zone but not in the lower monitoring zone. Therefore, it appears as if effluent has somehow bypassed the lower monitoring zone in the NE and NW parts of the site.

 DOI 10.1007/s10040-009-0511-6

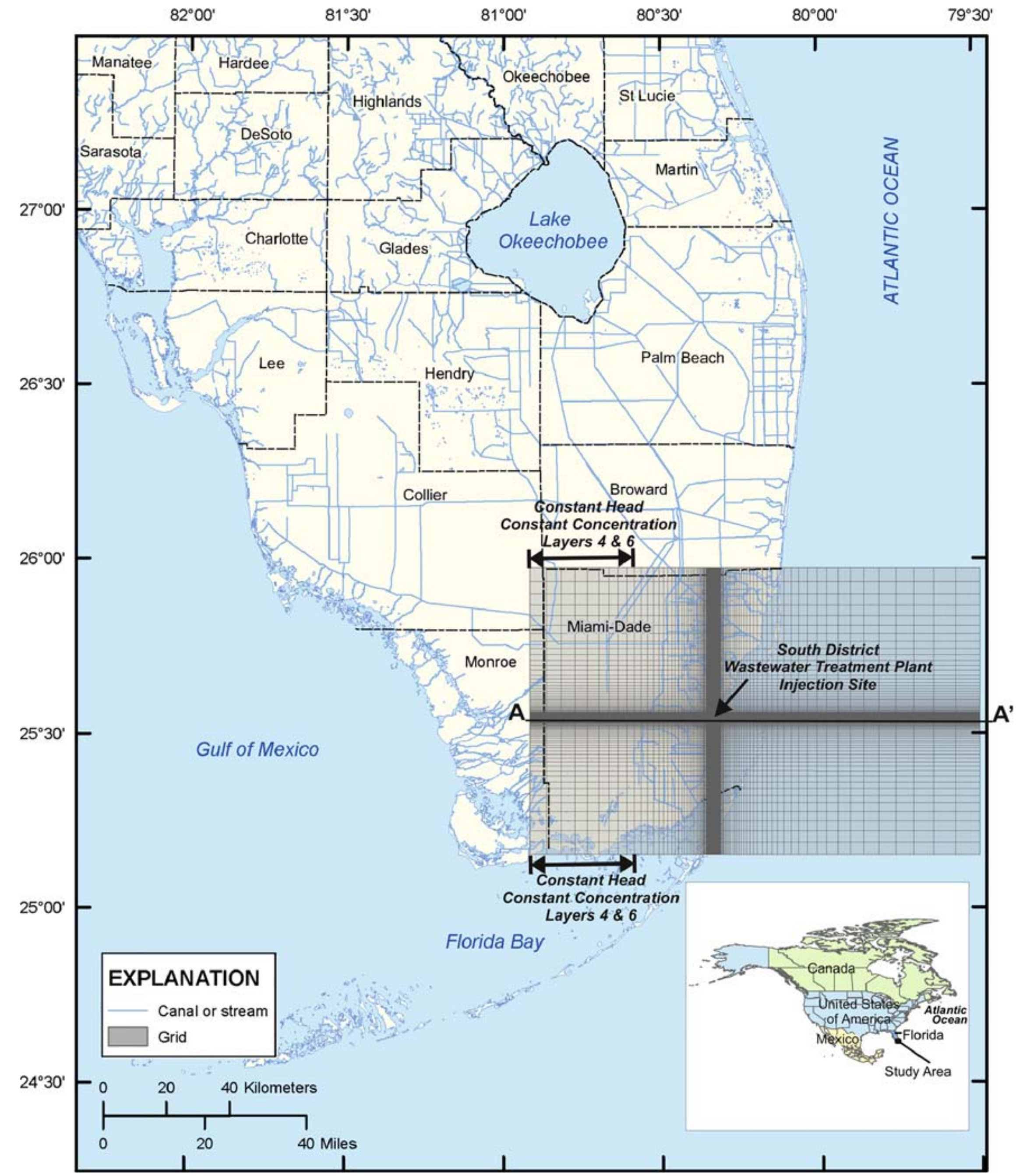

Fig. 1 Map of Florida showing the site location, model grid, landward boundary conditions in model layers 4 and 6, and counties (shown as *dashed lines*)

Hypothesis for effluent plume migration

Vertical migration is most likely caused by salinity-induced buoyancy effects caused by injection of the low-total dissolved solids (TDS; less than 1,000 mg/L), low density, effluent into the higher TDS (35,000 mg/L), higher density saline waters of the LFA. Vertical migration may also be partially affected by the increased hydraulic pressures caused by injection. While density variations and pressure increases are the obvious causes of upward migration, it is not straightforward to identify actual migration pathways. Evaluation of the monitoring well data suggests that multiple hypotheses could be used to explain plume locations. For example, there could be a leak in one of the injection or monitoring wells or voids may exist in the annular space between the casing and the borehole. There were multiple construction-related problems associated with the casing of IW-2, providing a potential vertical pathway in the SE part of the site. Also, an investigation determined a leak in the casing of the BZ monitoring well on the eastern part of the site, where the casing passed through the MFA; the lower monitoring zones of the BZ well were plugged in 1995. This BZ leak could explain the NE plume detected in monitoring wells

DOI 10.1007/s10040-009-0511-6

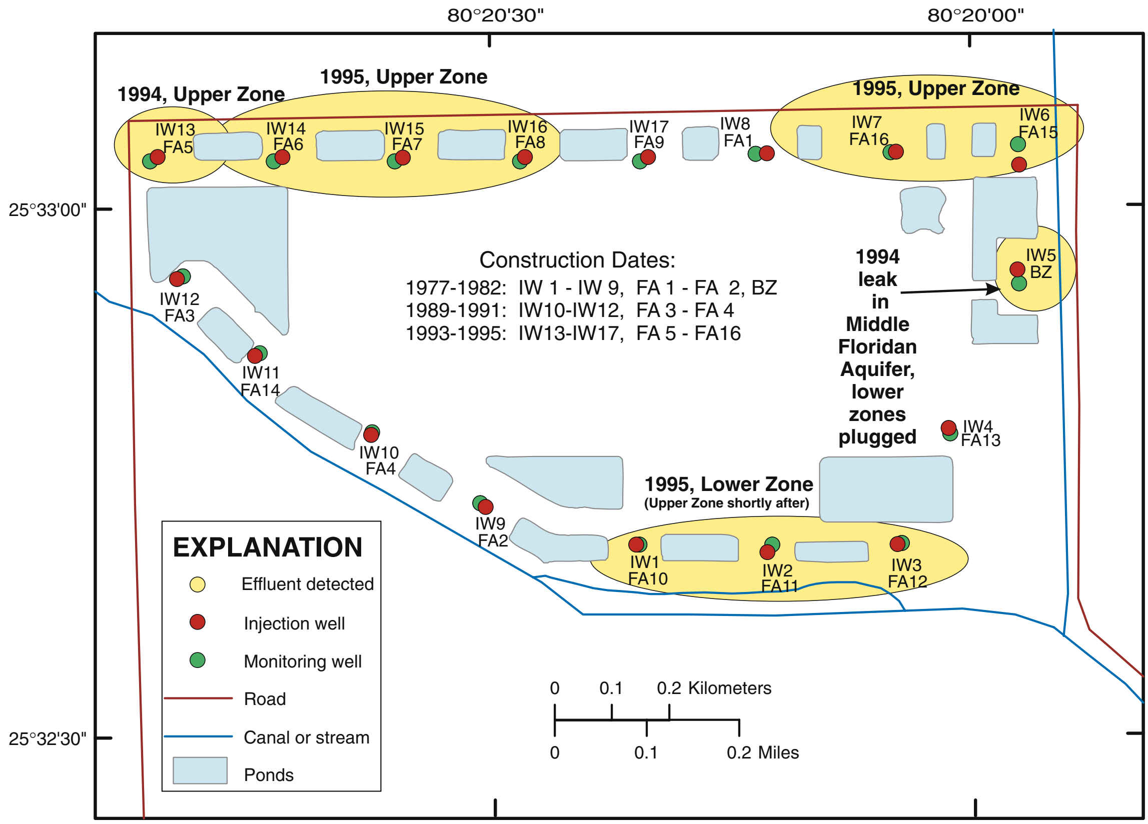

Fig. 2 Map of wastewater treatment plant showing injection (*IW*) and monitoring (*FA* and *BZ*) wells. Areas where effluent has been detected in a monitoring well are shown in *yellow*

FA-15 and FA-16 (these two wells did not monitor the LFA, Rust Environmental and Infrastructure 1998a, b). It has also been suggested that vertical pathways may be created during injection or monitoring well installation. Pilot holes with small bits are often drilled first to explore lithology. If subsequent reaming with a larger bit does not follow the pilot hole, vertical conduits may be introduced from the well installation.

Another hypothesis suggests that wells IW-8 to IW-17 were responsible for some effluent migration because these injection wells were not cased below the Delray Dolomite (McNeill 2000, 2002). The Delray Dolomite, a thin confining unit below the MCU2, was characterized using geophysical logs, temperature logs and lithologic drill cuttings at the SDWWTP, as well as cores from other regional sites. McNeill's work suggests that the Delray

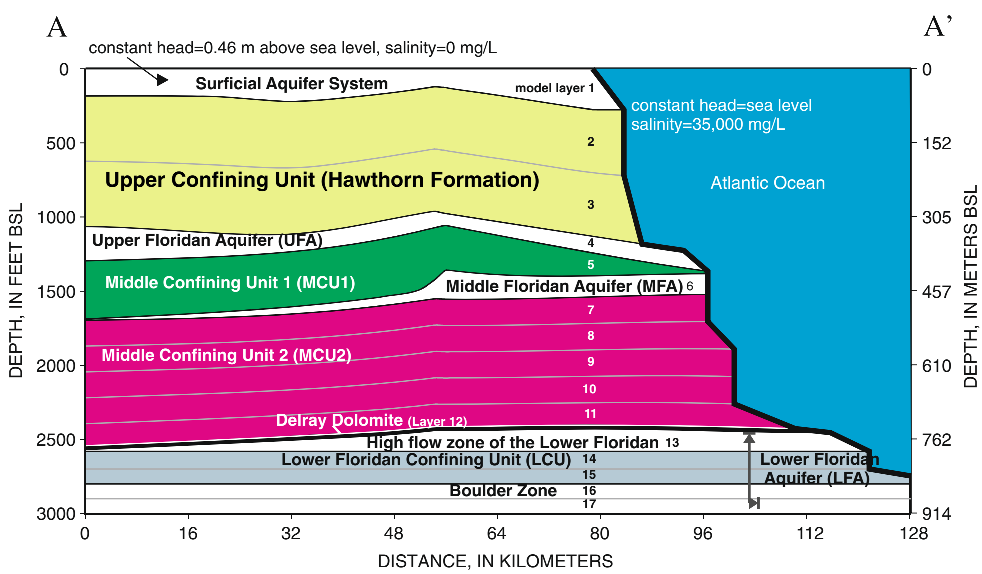

Fig. 3 Cross section showing aquifer and confining units (Reese and Richardson 2008; McNeill 2002). Model layers *1*–*17* are also shown

 DOI 10.1007/s10040-009-0511-6

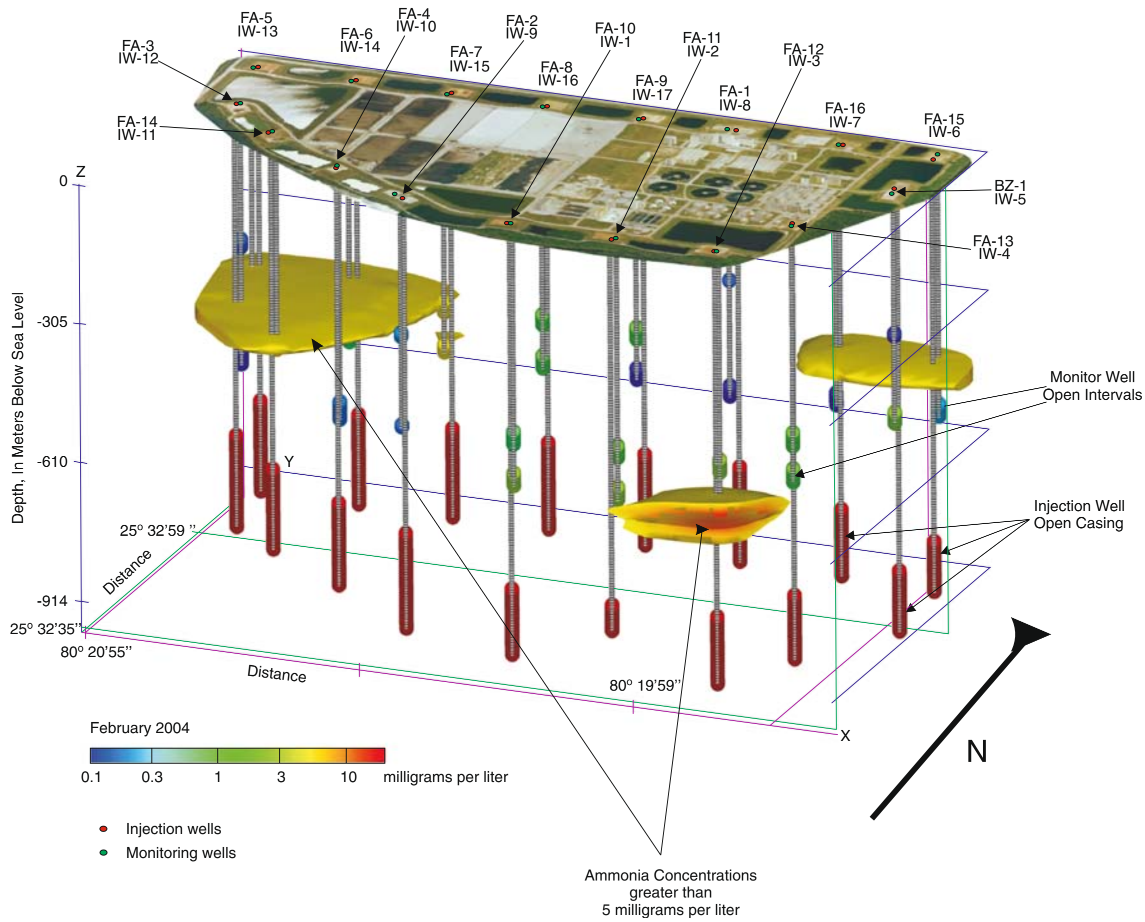

Fig. 4 Three-dimensional visualization of the site showing the injection and monitoring wells. The three effluent plumes are represented as *shaded areas* where ammonia concentrations exceed 5 mg/L. The NE and NW plumes are in the MFA; the SE plume is in the MCU2

Dolomite is an effective confining unit and prevents vertical migration of effluent. Based on hydraulic conductivities derived from aquifer tests and lab core tests, McNeill postulates that the Delray Dolomite is much more confining than the MCU2 (CH2M Hill 1977, 1981; MDWASD 1991; McNeill 2000, 2002). Alternatively, it is possible that the Delray Dolomite and MCU2 are only partially effective confining units (McNeill 2002) and that secondary porosity features, fractures, or other natural vertical conduits could be acting as a pathway for upward migration. Other interpretations of lithologic data at the site such as drill cuttings and down-hole video, suggest the presence of two faults in the Delray Dolomite and LFA (Winston 1995). Recent seismic work (K. Cunningham, U.S. Geological Survey, personal communication, 2008) has detected the presence of faults and other vertical features in the Floridan aquifer system, where it appears that vertical fractures could occur from as deep as 1,000 m below the land surface up to the land surface. Therefore, upward migration of effluent could occur through isolated vertical conduits or could occur as diffuse upward flow through continuous, but heterogeneous, confining layers.

Based on these hydrogeologic interpretations and an analysis of plume characteristics, the numerical model described in the next section is used to test the following hypotheses:

- Hypothesis 1—effluent plumes formed because of a construction problem at a well or due to a leak or break in a well casing.
- Hypothesis 2—effluent plumes formed because of a leaky confining unit or aquifer and aquitard heterogeneity.

Small and large scale heterogeneity can be equally important for describing plume migration (Carrera et al. 2005). While the geology of the study area is simple in general terms, comprising a sequence of relatively permeable and impermeable horizontally-oriented strata, factors affecting the movement of groundwater and effluent at the site scale are complex, with both geological and hydraulic conditions being important. Complexity may arise from any or all of the following factors.

1. There exists considerable potential for heterogeneity within the geological units comprising aquifers and within the aquitards that separate them. For example, Reese and Richardson (2008) describe the MCU1 and MCU2 as "semiconfining or leaky in nature and

 DOI 10.1007/s10040-009-0511-6

generally consisting of micritic limestone (wackestone to mudstone), dolomitic limestone, and dolomite or dolostone."

2. Vertical connectivity between the aquifers may be enhanced by local faulting and fracturing, the geometry of which is unknown.
3. Inter-aquifer connectivity induced during the construction of monitoring and injection wells or well failure may lead to effluent movement.

In light of these considerations, the focus of the modeling exercise was to test hypotheses that included well construction problems as well as geological heterogeneity as being responsible for the observed plume configurations. Hypotheses are tested through the calibration of a highly parameterized model in which hydrogeological heterogeneity can be introduced as a potential pathway for effluent migration.

Model development

A variable-density groundwater flow and solute transport model focused on the disposal site was developed using SEAWAT (a computer program for the simulation of three-dimensional variable-density groundwater flow and solute transport) Version 4 (Langevin et al. 2007). The SEAWAT model grid is comprised of 17 layers (Fig. 3), with 127 rows and 163 columns (Fig. 1). The model layers correspond to hydrostratigraphic layers as defined by Reese and Richardson (2008) and McNeill (2000, 2002). The horizontal discretization is more refined at the site, with cell sizes of 30 by 30 m, increasing gradually toward the model edge to a maximum cell size of 7,670 m. Although the model encompasses a rather large area compared to the site, this helps to ensure that simplifications made at model boundaries have minimal effects on effluent movement near the site. The model simulates a time period spanning January of 1983 through December of 2005 using an initial steady-state stress period, followed by 275 monthly transient stress periods. The model simulates the transport of two species, salinity and relative effluent concentration.

Boundary conditions

Constant-head cells representing the Atlantic Ocean are assigned to the model grid based on the intersection of sea floor bathymetry (Metzger and Buhmann 1993) with the model grid. The top layer of the model, representing the surficial aquifer system, also contains constant-head cells (Fig. 3). Immediately underlying the surficial aquifer system is the Hawthorne Formation, a thick semi-confining unit that hydraulically isolates the surficial aquifer from the Floridan aquifer system. Therefore, the constant-head cells representing the surficial aquifer system are likely to have little effect on effluent migration in the Floridan aquifer system. Below layer 1, only layers 4 and 6 contain constant-head boundaries on the landward side of the model (Fig. 1). Water levels and concentrations for these constant-head boundaries are estimated as part of the calibration process. Heads for the northernmost boundary in layers 4 and 6 are estimated as one parameter; heads for the southernmost boundary in layers 4 and 6 are estimated as a separate parameter.

Aquifer properties

Spatial variations of hydraulic conductivity were determined using the pilot point method (de Marsily et al. 1984). These variations were employed for both horizontal and vertical hydraulic conductivity in layers 4–12. As described by Doherty (2003), parameter values, in this case hydraulic conductivity, are estimated at locations within the model domain corresponding to those of the pilot points. Values are then assigned to individual model cells by interpolating between pilot points. In emplacing pilot points a philosophy of "the more the better" was used, with limits being set by computing resources. The use of a large number of pilot points prevents the occurrence of circular artifacts within the calibrated parameter field as the calibration process is provided the freedom to incorporate the information (such as potential heterogeneity) contained in the calibration data. Pilot points enable hypothesis testing to include geological heterogeneity as a possible mechanism for the migration of effluent.

Heterogeneity at both the site and regional scale was treated by using two sets of pilot points in layers 4–12. One set of pilot points with high spatial density is used to represent hydraulic conductivity at the site (these are referred to as the "inner" pilot points), while a second set of points of lower spatial density is used to represent hydraulic conductivity away from the site (these being referred to as "outer" pilot points). The use of a lower density away from the site allows the total number of pilot points to be reduced while maintaining flexibility to accommodate heterogeneity at different scales. Spatial interpolation between the outer points (numbering 36 in each layer) is accomplished using ordinary kriging based on a log transformed exponential variogram with a range of 90 km. The set of 158 "inner" pilot points are spaced much closer together at the site to enable the introduction of high spatial variability of hydraulic conductivity in the area surrounding the injection wells if necessary. Parameters associated with the inner pilot points are log-transformed hydraulic conductivity multipliers of the conductivity field obtained by interpolation from outer pilot points. Spatial interpolation between the inner points is accomplished using simple kriging based on an exponential variogram with a range of approximately 457 m and a mean value of 1.0. This, in conjunction with the Tikhonov regularization scheme (Tonkin and Doherty 2005) described later, ensures that multiplier parameters estimated for inner points and multiplier values at model cells interpolated from these points deviate minimally from 1.0 only if the data guides the calibration process to do so. The use of this dual inner-outer parameterization

 DOI 10.1007/s10040-009-0511-6

(with inner parameters constituting multipliers for outer parameters) allows the smooth transition of hydraulic conductivity from the site scale, where greater spatial variability of hydraulic properties is allowed, to the regional scale where less spatial variability is necessary.

The hydraulic conductivity of other model layers (Surficial, Hawthorn, Boulder Zone LFA and Lower Confining Unit), as well as porosity and storage in all 17 layers, is assumed to be spatially uniform and is represented by a single parameter in each case. Other estimated parameters (included or excluded according to the hypotheses tested) include conductance values pertaining to two Drain Return Flow, DRT, (Banta 2000) cells. The DRT package enables effluent to be removed from one cell and discharged into another, and is used here to represent a leak within a borehole or well. At other locations within the model domain, vertical columns of one cell width spanning layers 6–12 were assigned high values of vertical hydraulic conductivity to allow simulation of effluent movement along well casings (referred to herein as "chimneys"). Both the vertical hydraulic conductivities and porosities of these "chimneys" were estimated during the hypothesis-testing calibration processes. All parameters estimated through the regularized calibration process are listed in Table 1.

Initial conditions

Initial heads and concentrations for density-dependent groundwater models are typically obtained by running the model for a long period of time under a pre-development stress regime (Misut and Voss 2007). Unfortunately, this is a computationally intensive procedure, particularly for deep aquifers with slow response times, such as the Floridan aquifer system (Hughes et al. 2007). The present model takes approximately 75 min to simulate effluent movement over 23 years; hence a few days would be required to compute initial conditions using this approach. Additionally, because initial conditions depend on values assigned to the model parameters, each parameter estimation run would require an initial condition run; overall calibration times would thus be prohibitive.

In some of the preliminary simulations, modeled salinities and heads did not match measured values well, particularly for the period prior to injection. This disagreement, which was primarily due to poor initial conditions, was problematic for estimating parameters because the calibration process adjusted parameters to compensate for the poor initial conditions. To mitigate this problem, initial head and salinity conditions that agreed with measured values prior to injection were calculated using a semi-analytical approach; pertinent initial condition parameters were then estimated as part of each calibration process. Characterization of the salinity field was reduced from three dimensions to two by matching the depth to the interface (the "depth" is assumed to be at 50% freshwater and 50% seawater, or a TDS of 17,500 mg/L) to measured field observations, and then projecting this surface onto the three-dimensional grid. The transition zone from freshwater to seawater is represented using an analytically characterized continuous-concentration profile with a transition zone width estimated through model calibration. Initial salinities above this transition zone were assigned a freshwater value of 0 mg/L while salinities below the transition zone were assigned seawater salinities of 35,000 mg/L—for more detail on the method, see the ELEV2CONC utility (used to compute the elevation of the freshwater/saltwater interface on the basis of a sequence of concentration arrays) contained in the PEST (Model-Independent Parameter Estimation and Uncertainty Analysis) Groundwater Data Utilities Suite from Doherty (2007). After salinities were estimated, one steady-state forward run of the model provided simulated water levels, which were then compared to pre-injection measured water levels.

Estimating initial heads and concentrations as part of the overall parameter estimation process in this manner provides the following advantages: (1) initial conditions are hydraulically realistic and are consistent with the model's current parameter field; (2) initial heads and salinities are consistent with measured values; and (3) computer runtimes are substantially shorter than run times for alternative approaches that require running the model to equilibrium.

Calibration procedure

For the type of hypothesis testing required in this study, the calibration process must possess the capacity to introduce geologically reasonable heterogeneity at locations within the model domain that are not necessarily known before the testing process is instigated. Hence the calibration process must accommodate the use and estimation of a large number of parameters. Estimation of these parameters must therefore take place using mathematical regularization methodologies which are designed specifically for use in such calibration contexts.

Two types of mathematical regularization were employed in the present study, both of these being implemented in the PEST calibration package (Doherty 2009). Tikhonov regularization provides a calibration framework in which parameter values, or relationships between parameter values, are preferentially maintained at user-specified levels which are considered to be of maximum likelihood in the prevailing geological context. Perturbations from these values are introduced only if necessary, only where necessary, and only to the minimal amount necessary to achieve an acceptable level of fit between model outputs and field measurements. At the same time over-fitting is prevented through limiting goodness of fit beyond a user-specified level. In the present study, Tikhonov constraints were comprised of preferred values for all parameters — including hydraulic conductivities or hydraulic conductivity multipliers as previously described.

In order to maintain unconditional numerical stability of the highly parameterized calibration process, and in order to accelerate that process, subspace regularization, in

 DOI 10.1007/s10040-009-0511-6

Table 1 Parameters estimated as part of the hypothesis testing calibration process

Estimated parameter(s)	Parameterization methodology	Number of parameters used
Vertical/horizontal hydraulic conductivity for layers 1–3, 13	Uniform	3
Horizontal hydraulic conductivity for layers 14–17	Uniform	2
Vertical hydraulic conductivity for layers 14–17	Uniform	2
Vertical/horizontal hydraulic conductivity for layers 4–5, 7–12	Pilot points	1,164
Horizontal hydraulic conductivity for layer 6	Pilot points	194
Vertical hydraulic conductivity for layer 6	Pilot points	194
Porosity for layers 1–17	Uniform	17
Storage for layers 1–17	Uniform	17
Drain return flow (DRT) conductance	Single value	2
Chimney-hydraulic conductivity	Single value	1
Chimney-porosity	Single value	1
Water levels-constant-head cells along north in first 9 columns in layers 4 and 6	Uniform	1
Water levels-constant-head cells along south in first 9 columns in layers 4 and 6	Uniform	1
Depth to interface (at three wells)	Single value	3
Interface thickness	Single value	1

the form of singular value decomposition, was also introduced to the inversion process. Prior to the commencement of any of the calibration-based hypothesis testing procedures already discussed, singular value decomposition was employed to define orthogonal parameter combinations (referred to as "super parameters" in PEST parlance), that could be estimated in lieu of the parameters themselves; this procedure is fully described in Tonkin and Doherty (2005). Because the number of super parameters to be estimated is normally considerably smaller than the number of native parameters employed by a model, the numerical burden of parameter estimation is reduced considerably through adoption of this strategy. A total of 230 super parameters were employed in calibration of the current model, this number coinciding with the number of computing nodes available for parallel calculation of parameter sensitivities during each iteration of the calibration process (see the following). Information on super parameter sensitivities forthcoming from the calibration process indicated that only about 100 of these were actually estimated during any one iteration of the calibration process (the number changed somewhat from iteration to iteration as a result of model nonlinearity). Numerical stability of the estimation procedure was maintained through the method of hybrid Tikhonov-subspace inversion; estimating only as many super parameters as were locally adjustable using a truncated singular value decomposition solution procedure, this being reinforced by the Tikhonov procedure previously discussed.

A single forward model runs for approximately 75 min on a server with a 64-bit pathway, a 500 GB hard drive, a 2.66 GHz processor, and 16 GB of memory. Hence approximately 2,000 hours are required to calculate one Jacobian matrix using finite-parameter differences. To reduce the simulation time, runs were parallelized across a Microsoft Windows cluster comprised of 232 processor cores, thus reducing calculation time to just under 9 hours. Once calculation of the Jacobian matrix was complete, super parameters were defined; these are then estimated (subject to Tikhonov constraints on native model parameters) through the ensuing calibration process. 230 super parameters were chosen for estimation in order to maximize the use of these computing resources, and truncated singular value decomposition on these super parameters was used as an additional constraint.

Using this approach, more than one hypothesis could be tested in a relatively short amount of time. Each such test required attempted re-calibration of the model subject to boundary conditions and construction details pertinent to a particular hypothesis. Failure to match historical observations of contaminant movement through this process, or achievement of a match only through the use of unrealistic parameters, indicates an invalid hypothesis. In this particular problem, the concept of unrealistic parameter values is difficult to define because of the potential for fracture flow in the carbonate system. Therefore, parameter values could be much higher than suggested in literature that fails to account for the prospective fracture flow.

Results

Three different calibration runs were performed to test the hypotheses regarding plume migration pathways. The three calibration runs are as follows:

Calibration run 1. All three plumes result from problems in well construction. This is a test of hypothesis 1 for all plumes. The SE plume results from well construction problems at IW-2; the NE plume is caused by a leak in the BZ monitoring well; the NW plume is caused by a leak in FA-6 (Fig. 2).

Calibration run 2. The NW plume is caused by a leaky confining unit. For this calibration run, the short circuit at FA-6 is removed. Thus, the calibration procedure is forced to introduce heterogeneity in aquifer properties as the sole

 DOI 10.1007/s10040-009-0511-6

mechanism for providing a pathway for effluent to migrate vertically, testing hypothesis 2 for the NW plume. The short circuit pathways at BZ and IW-2 are retained in this calibration run.

Calibration run 3. The SE plume is caused by a leaky confining unit. For this calibration run, the short circuit at IW-2 is removed. The calibration procedure is forced to introduce heterogeneity in aquifer properties in this area as the sole mechanism for effluent to migrate vertically, testing hypothesis 2 for the SE plume. The short circuit pathways at BZ and FA-6 are retained in this calibration run.

Because the leak in the BZ monitoring well was known to have occurred, it was active in all calibration runs until 1995 (when the leak was plugged); therefore the leaky confining unit hypothesis was not tested for the NE plume as this plume appears to be caused by the BZ leak.

Acceptance or rejection of a hypothesis is first determined by comparing observed data with simulation results. If the match is poor, then the hypothesis is rejected. If the calibrated model is able to match the historical plume development, then the resulting parameter fields are closely evaluated to determine if they are reasonable or not. If the calibrated model can reproduce the historical plume development patterns and the estimated parameters are reasonable, then the hypothesis is retained. Those calibrated models that match observations and have reasonable parameter fields can then be used to predict future plume migration patterns.

Northwest plume

Hypothesis 1. is tested with the first calibration run for the NW plume. Calibration results for FA-6 (Fig. 5a–c) indicate that a leak at FA-6 is a possible explanation for the plume development. Simulated water levels match relatively well with measured water levels (Fig. 5a). Seasonal changes in injection rates do appear to affect the measured water levels, revealing that the entire transient nature of the system is not completely captured by hypothesis 1. Even though simulated water levels do not exhibit the same seasonal temporal variability as their measured counterparts in Fig. 5a, the overall long-term transient nature of the system appears to be captured by hypothesis 1 in the model. Simulated salinity and effluent fractions for FA-6 upper and lower are also in agreement with field observations (Fig. 5b, c). A spike in effluent fraction is seen in FA-6 upper and the increase in effluent is more rapid than that seen in the field; however, the general trend is captured. The imposed construction problem enables the effluent to bypass the lower monitoring well and appear in the upper monitoring well.

Hypothesis 2. is tested with the second calibration run. The model cannot match observed concentrations using only a leaky confining unit (Fig. 5a–c). Even though the water levels match well, the inability of the model to allow the effluent to migrate to the NW area of the site suggests that a leaky confining unit is an unlikely mechanism for the NW plume. Despite the flexibility provided in the model to represent areas of high hydraulic conductivity, effluent migration to the NW does not occur.

The results from calibration graphs demonstrate that the NW plume is potentially caused from a construction related problem such as a leaky casing or a misaligned borehole. This analysis of the NW plume indicates that hypothesis 1 can be retained barring evaluation of parameter values and estimated parameter fields, whereas hypothesis 2 must be rejected.

Southeast plume

Hypothesis 1. is tested with the first calibration run for the SE plume. Results reveal that model calibration based on hypothesis 1 for the SE plume appears possible (Fig. 5d–f). Similar to the results for the NW plume, the simulated water levels do not appear to capture the short-term transient nature of the system; however, the long-term transient nature of the system does appear to be simulated by the model. Measured and simulated salinity concentrations for hypothesis 1 match well, providing additional support for hypothesis 1 for the SE plume.

Hypothesis 2. is tested with the third calibration run for the SE plume and also appears valid according to Fig. 5d–f. Simulated concentrations match very well to their measured counterparts (Fig. 5e, f). Also, hypothesis 2 does appear to capture the short and long-term transient nature of the system (Fig. 5d) where simulated water levels exhibit the same seasonal temporal variability as their measured counterparts.

The resulting calibration graphs demonstrate that neither hypothesis 1 nor 2 pertaining to the origin of the SE plume can be rejected; therefore, these hypotheses are retained barring evaluation of parameter values and estimated parameter fields. Effluent migration from a construction related problem or from leaky confining units are both plausible at this stage.

Estimated parameters

To accept or reject hypothesis 1 for the three plumes or hypothesis 2 for the SE plume, estimated pilot point parameters and hydraulic conductivity fields are also analyzed for the retained calibration runs that test these hypotheses (calibration runs 1 and 3 were retained, whereas calibration run 2 has been rejected). The model estimates for the pilot point parameters are all within the range of published estimates with the exception of the Delray Dolomite (Table 2) and the low-end estimates of the MFA. The estimated vertical hydraulic conductivity of the Delray Dolomite in the retained calibrated models is between 0.019–0.4 m/day, which is higher than predicted by McNeill (2000, 2002), where the maximum estimate for the Delray Dolomite is ~0.0025 m/day (Table 2). However, the Delray Dolomite is not considered to be a

 DOI 10.1007/s10040-009-0511-6

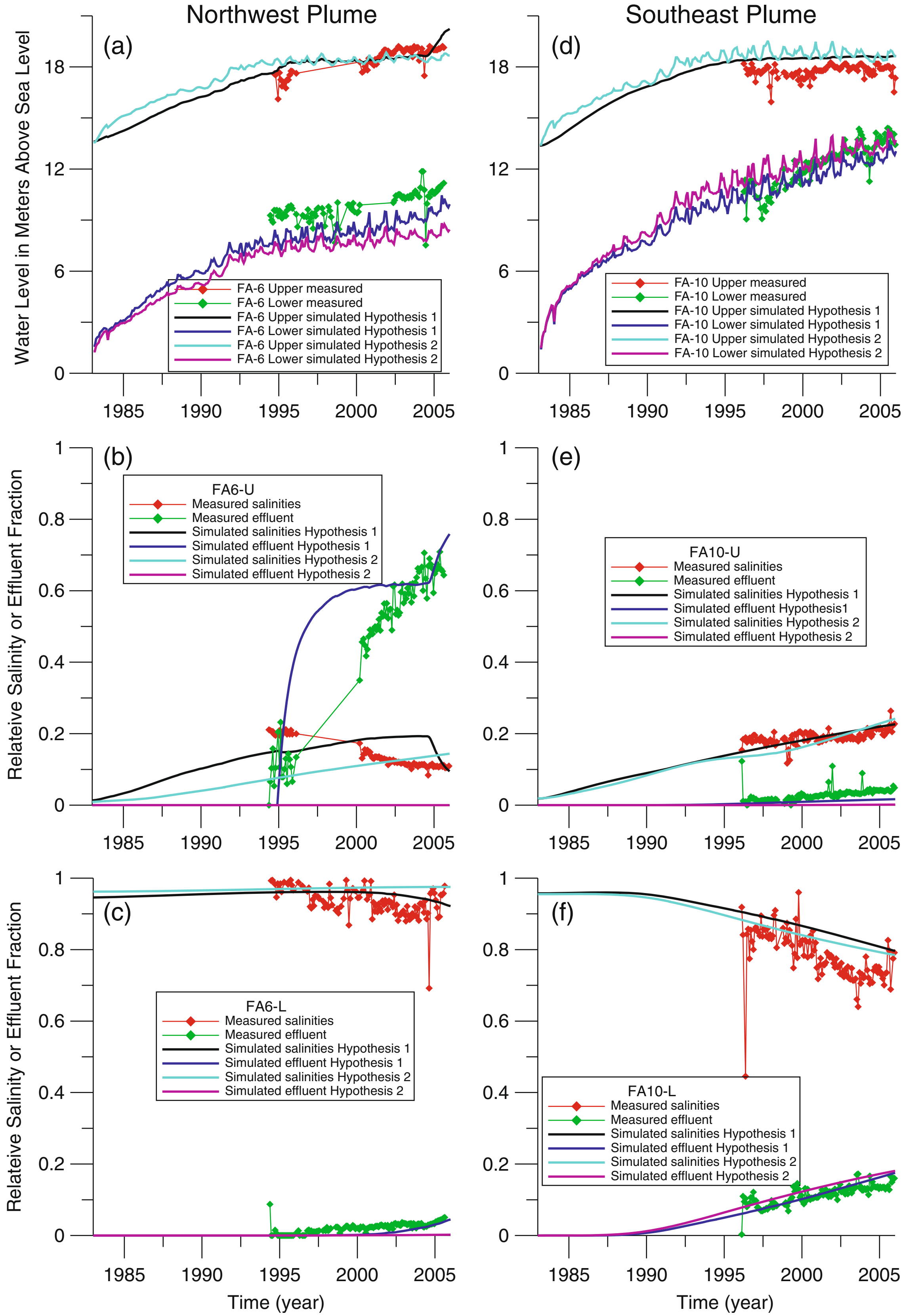

Fig. 5 Calibration graphs for **a** water levels at FA-6, **b** relative salinity and effluent fractions for FA-6 upper, **c** relative salinity and effluent fractions for FA-6 lower, **d** water levels at FA-10, **e** relative salinity and effluent fractions for FA-10 upper, and **f** relative salinity and effluent fractions for FA-10 lower

DOI 10.1007/s10040-009-0511-6

Table 2 Range of transmissivities for aquifers and vertical hydraulic conductivities for aquitards according to published literature (Meyer 1974; Reese and Richardson 2008; McNeill 2000, 2002) and the estimated values from the pilot point parameters in calibration runs 1 and 3. (Note: MCU1 and MCU2 are combined because the published literature combines the two)

Pilot point parameter	Range of transmissivity or hydraulic conductivity from literature	Minimum value for calibrated models	Maximum value for calibrated models
UFA (transmissivity)	~12–28,000 m^2/day	22 m^2/day	1,670 m^2/day
MFA (transmissivity)	~12–150,000 m^2/day	3 m^2/day	4,128 m^2/day
MCU1/MCU2 (vertical hydraulic conductivity)	~9E-6–8.5 m/day	5.5E–6 m/day	6.5 m/day
Delray dolomite (vertical hydraulic conductivity)	~9E-6–2.5E-3 m/day	0.019 m/day	0.4 m/day

separate layer according to Reese and Richardson (2008), where the Delray Dolomite is considered part of MCU2 and vertical estimates of hydraulic conductivity could, therefore, be higher (Table 2). In this case, the model-estimated vertical hydraulic conductivities of the Delray Dolomite are within the range of reported literature values for MCU2. The estimates of hydraulic conductivity for the MFA are typically within the range of literature values, with a few of the pilot point values falling below the low-end published values (Table 2).

When looking at the estimated hydraulic conductivity fields in the Delray Dolomite for the retained calibration runs (Fig. 6a, b), the results of these two calibration exercises could support both hypotheses for contaminant migration. The construction problems associated with IW-2 and in the MFA may indeed provide a migration pathway for effluent according to hypothesis 1 (Fig. 6a). Alternatively, enhanced hydraulic conductivity resulting, for example, from local faulting (Winston 1995) could also be responsible for the SE plume according to hypothesis 2 (Fig. 6b). It is also possible that the hydraulic conductivity field in Fig. 6b could result from fracturing of the Delray Dolomite in response to excessive drilling at the site. Yet another possibility is that since IW-8 to IW-17 were drilled and left open below the Delray Dolomite, but cased above it, there may be residual holes left in the Delray Dolomite layer where vertical migration could occur through the Delray Dolomite. However, the estimated hydraulic conductivity in Fig. 6a is notably higher than that estimated across the rest of the model, which is more difficult to explain geologically than leakage along a well casing. Thus hypothesis 1 seems more likely than hypothesis 2 for the SE plume after analyzing the calibrated hydraulic conductivity fields and researching the site history. At the same time, neither hypothesis can be rejected with certainty.

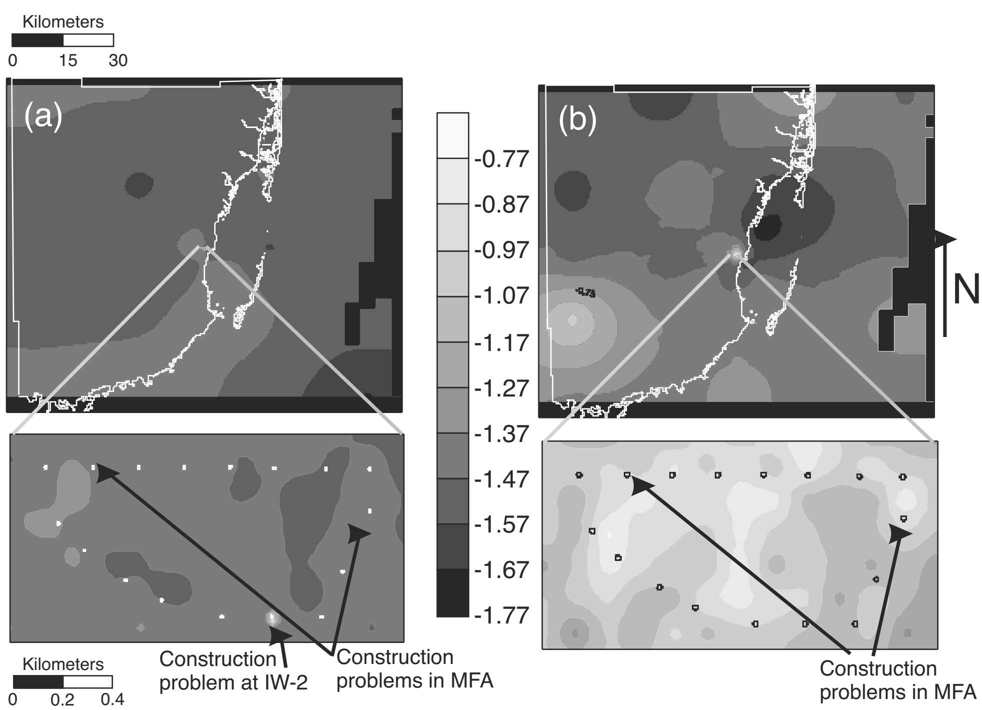

Fig. 6 Calibrated log transformed vertical hydraulic conductivity fields (in m/day) for the Delray Dolomite (Fig. 3) for **a** calibration run 1 with construction problems for all three plumes, and **b** calibration run 3 which tests hypothesis 2 for the SE plume. Injection wells are the *small white or black open circles*

 DOI 10.1007/s10040-009-0511-6

Discussion and conclusions

This paper has described a modeling approach, which departs from that which is commonly used in environmental management in the ways described in the following. Firstly, the uncertainty with which predictions of interest can be made is acknowledged, and explicitly accommodated, as part of the modeling process by developing one model and then using that model to test different hypotheses through sequential calibration exercises against the same set of historical field data—this is recommended by Beven and Binley 1992; Beven and Freer 2001; Poeter and Anderson 2004; Ye et al. 2005, 2008; and by Carrera et al. 2005.

Secondly, despite the large model run-times, hypothesis testing is still based on highly parameterized inversion. Tractability of the calibration process is achieved through a high level of model run parallelization and through the method of hybrid Tikhonov-subspace inversion that requires estimation of a number of super parameters, which are much less than the actual number of parameters used in the model.

The challenge facing most calibration exercises is to extract as much information as possible from the calibration dataset and to transfer that information to the model. This challenge can be especially salient when the calibration process becomes one of hypothesis-testing. Rejection of a hypothesis because of (1) lack of fit with historical data, or (2) a fit resulting in the use of unlikely parameters or parameter fields, should only take place after the calibration process has been given the opportunity to prevent it. Highly parameterized inversion can provide a model with maximum flexibility for the existence of complex phenomena such as localized or pervasive heterogeneity of hydraulic properties. This is particularly important in the present study where observations showed movement of a previously unpredicted contaminant through a carbonate system. However numerical stability, computational tractability, and minimization of the chances of incurring unrealistic parameter values in the calibration process is made possible with the undertaking of highly parameterized inversion using efficient regularization methods.

The present study has demonstrated that the explanation for vertical migration of effluent to upper geological layers could be simple—an outcome of poor borehole construction. It has also demonstrated that its vertical migration to upper geological layers could result from a higher level of hydrogeological heterogeneity in and above the aquifers used for waste disposal than was originally assumed; however, heterogeneity cannot explain all the effluent migration. A map view of effluent disposition in the MFA for all three calibration runs is shown in Fig. 7a–c. Both of the effluent distributions in Fig. 7a, c are consistent with what is currently known of the site and are supported by the data, whereas the effluent distribution in Fig. 7b is not supported by the data. Though there are broad similarities between the effluent distribution in calibration runs 1 and 3 (Fig. 7a, c), there are some differences, as seen in the western portion of these depictions, where the NW plume is located (Fig. 7a, c). When analyzing the plumes in this manner, it is important to note that future predictions based on multiple hypotheses (that are supported by the data) may provide different results. In demonstrating mechanisms for contaminant movement in this fashion, the hypothesis testing procedure implemented through regularized inversion has also suggested where extra information or data could be gathered which may allow confirmation (or possibly rejection) of prevailing hypotheses.

The present study has presented unique challenges. Included among these challenges was the need to rapidly re-compute initial head and concentration fields which are compatible with parameter sets that are continuously altered through the calibration process. This was accomplished using a semi-analytical description of initial salt concentrations within the region of the freshwater–saltwater interface, and by adjusting the parameters of this semi-analytical equation and the hydraulic property parameters employed by the model as part of the calibration process.

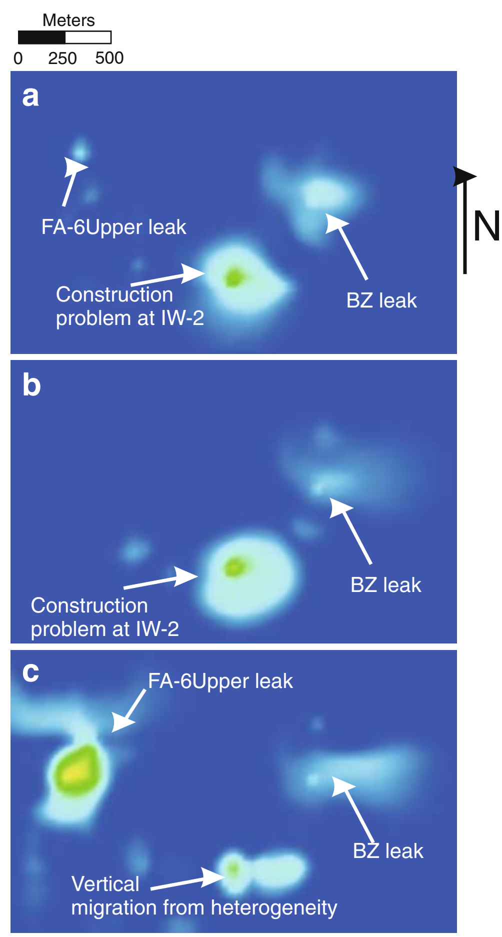

Fig. 7 Map view of effluent in the MFA (Fig. 3) showing the migration of effluent from **a** calibration run 1 (which represents hypothesis 1 for all plumes), **b** calibration run 2 (which represents hypothesis 2 for the NW plume), and **c** calibration run 3 (which represents hypothesis 2 for the SE plume)

 DOI 10.1007/s10040-009-0511-6

In most real-world modeling exercises, there is an attempt to extract as much information as possible from an historical (and often expensive) dataset, irrespective of the purpose of the model. Flexibility in parameterization provides the calibration process maximum freedom in either challenging the current conceptual model, or of confirming that this conceptual model is compatible with historical observations. As described by Hunt et al. (2007), regularized inversion allows the calibration process to undertake parameter simplification to the extent that this is required, while simultaneously retaining parameterization detail where this is supported by the calibration dataset.

This study also demonstrates that the implementation of high-end model calibration through highly parameterized inversion is computationally feasible in modern computing environments. This is largely due to the parallelizability of model runs, which is a major time-saving necessity for this process. The hybrid Tikhonov-subspace singular value decomposition scheme that estimates a number of super parameters, which are considerably smaller in number than native parameters, lends tremendous support in reducing the computational burden of this methodology. The shortcut of estimating initial conditions further supports the reduction in calibration times.

The site presented here is complicated, with many possibilities—from problems with well construction to diffuse upward flow through semi-confining units to aquifer and aquitard heterogeneity—for the vertical migration of effluent. Hypothesis testing using a highly parameterized model appears to have successfully provided the flexibility needed to explore the likelihood associated with multiple reasons for the unexpected vertical migration of effluent. The numerical modeling here shows that one hypothesis can be accepted, while another can be rejected. Construction related issues may be more likely than hydrogeologic heterogeneity for some of the vertical migration, particularly in the areas where effluent is observed in upper monitoring zones, but not in the adjacent lower monitoring zones. However, in some instances, heterogeneity is also a possibility. Each plausible mechanism provides a different picture of the current effluent disposition between monitoring points. Calibrated models achieved through the hypothesis testing process described previously, each based on a different contaminant migration mechanism, will be run into the future in order to provide different possibilities for the future disposition of the contaminant in the upper geological layers.

Acknowledgements The authors are grateful for the reviews by L. Brakefield-Goswami, M. Tonkin, S. Hunter, S. Duncan, and the reviewers selected by *Hydrogeology Journal*. Gratitude is also extended to Guest Editors E. Abarca and V. Post for selecting this research to be a part of a special issue. This work was partially funded by the United States Geological Survey and Miami Dade Water and Sewer Department. Virginia Walsh at Miami Dade County was integral in the completion of this work.

References

Banta ER (2000) MODFLOW-2000, the U.S. Geological Survey modular ground-water model: documentation of packages for simulating evapotranspiration with a segmented function (ETS1) and drains with return flow (DRT1). US Geol Surv Open-File Rep 127 pp

Beven KJ, Binley AM (1992) The future of distributed models: model calibration and uncertainty prediction. Hydrol Process 6 (3):279–298

Beven KJ, Freer J (2001) Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. J Hydrol 249:11–29

Bredehoeft J (2005) The conceptualization model problem: surprise. Hydrogeol J 13:37–46

Carrera J, Alcolea A, Medina A, Hidalgo J, Slooten LJ (2005) Inverse problem in hydrogeology. Hydrogeol J 13:206–222

CH2M Hill (1977) Drilling and testing of the test-injection well for the Miami-Dade Water and Sewer Authority. Contract S-153, Miami-Dade Water and Sewer Authority, Miami

CH2M Hill (1981) Drilling and testing of the 8 injection wells and the 3 monitoring wells for the South District Regional Wastewater Treatment Plant of the Miami-Dade Water and Sewer Authority. Contract S-154, Miami-Dade Water and Sewer Authority, Miami

de Marsily GH, Lavedan G, Boucher M, Fasanino G (1984) Interpretation of interference tests in a well field using geostatistical techniques to fit the permeability distribution in a reservoir model. In: Verly et al (ed) Proc Geostatistics for natural resources characterization. Part 2. D. Reidel, Dordecht, The Netherlands, pp 831–849

Doherty J (2003) Groundwater model calibration using pilot points and regularization. Ground Water 41(2):170–177

Doherty J (2007) Groundwater data utilities, part B: program descriptions. Watermark, Brisbane, Australia

Doherty J (2009) PEST: model-independent parameter estimation user manual. Watermark, Brisbane, Australia

Hughes JD, Vacher HL, Sanford WE (2007) Three-dimensional flow in the Florida platform: Theoretical analysis of Kohout convection at its type locality. Geology 35(7):663–666

Hunt RJ, Doherty J, Tonkin MJ (2007) Are models too simple? Arguments for increased parameterization. Ground Water 45 (3):254–262

Jamieson GR, Freeze RA (1983) Determining hydraulic conductivity distributions in a mountainous area using mathematical modeling. Ground Water 21(2):168–177

Krabbenhoft DP, Anderson MP (2006) Use of a numerical ground-water flow model for hypothesis testing. Ground Water 24 (1):49–55

Langevin CD, Thorne D, Dausman AM, Sukop MC, Guo W (2007) SEAWAT Version 4: a computer program for simulation of multi-species solute and heat transport. US Geol Surv Tech Methods Book 6, Chapter A22, 39 pp

Maliva RG, Guo W, Missimer T (2007) Vertical migration of municipal wastewater in deep injection well systems, South Florida, USA. Hydrogeol J 15(7):1387–1396

McNeill DF (2000) Final report: a review of upward migration of effluent related to subsurface injection at Miami-Dade water and Sewer South District Plant. McNeill Geological Services, Miami, FL

McNeill DF (2002) A geological review of the confining capability of a regional dolomite unit: application to the MDWAS south district WWTP. Prepared for the Miami-Dade County Attorney, McNeill Geological Services, Miami

Metzger DR, Buhmann RH (1993) National Environmental Satellite, Data, and Information Service. http://www.ngdc.noaa.gov/mgg/dat/geodas/docs/aro88.htm. 5 August 2009

Meyer FW (1974) Evaluation of hydraulic characteristics of a deep artesian aquifer from natural water-level fluctuations, Miami, Florida. Report of Investigations 75, Bureau of Geology, Tallahassee, FL, 32 pp

DOI 10.1007/s10040-009-0511-6

Miami Dade Water and Sewer Department (MDWASD) (1991) Drilling and testing of 3 injection wells and 2 monitoring wells for the South District Regional Wastewater Treatment Plant. Contract S-330, MDWASD, Miami

Misut PE, Voss CI (2007) Freshwater-saltwater transition zone movement during aquifer storage and recovery cycles in Brooklyn and Queens, New York City, USA. J Hydrol 337:87–103

Pennequin DFE (1983) Ground-water flow model determinations of area hydrogeology and geology. Ground Water 21(5):552–557

Poeter E, Anderson D (2004) Multimodel ranking and inference in ground water modeling. Ground Water 43(4):597–605

Reese RS, Richardson E (2008) Synthesis of the hydrogeologic framework of the Floridan Aquifer System and delineation of a major Avon Park permeable zone in central and southern Florida. US Geol Surv Sci Invest Rep 2007-5207

Remson I, Gorelick SM, Fliegner KF (1980) Computer models in ground-water exploration. Ground Water 18(5):447–451

Rust Environmental and Infrastructure, Inc. (1998a) Report No. 1: chemical tracer feasibility study, South District Wastewater Treatment Plant, Dade County, Florida. Rust Environmental and Infrastructure, Miami

Rust Environmental and Infrastructure Inc. (1998b) Report No. 1: geochemical study report for South District Wastewater Treatment Plant, Dade County, Florida. Rust Environmental and Infrastructure, Miami, FL

Tonkin M, Doherty J (2005) A hybrid regularized inversion methodology for highly parameterized models. Water Resour Res 41, W10412. doi:10.1029/2005WR003995

Troldborg L, Refsgaard JC, Jensen KH, Engesgaard P (2007) The importance of alternative conceptual models for simulation of concentrations in a multi-aquifer system. Hydrogeol J 15:843–860

Winston GO (1995) The boulder zone dolomites of Florida, vol 1, Paleogene and upper cretaceous zones of the southeastern peninsula and the Keys. Miami Geological Society, Miami

Ye M, Neuman SP, Meyer PD, Pohlmann K (2005) Sensitivity analysis and assessment of prior model probabilities in MLBMA with application to unsaturated fractured tuff. Water Resour Res 41, W12429. doi:10.1029/2005WR004260

Ye M, Meyer PD, Neuman SP (2008) On model selection criteria in multimodel analyses. Water Resour Res 44, W03428. doi:10.1029/2008WR006803

 DOI 10.1007/s10040-009-0511-6