The importance of alternative conceptual models for simulation of concentrations in a multi-aquifer system

Lars Troldborg · Jens Christian Refsgaard · Karsten Høgh Jensen · Peter Engesgaard

Abstract Four different conceptual models based on alternative geological interpretations were formulated for a shallow 600km^2 aquifer system in Denmark comprising Quaternary deposits. Each of the four models was calibrated against groundwater heads and discharge measurements through inverse modeling. Subsequently, the transport capabilities of the four models were compared to 32 concentration measurements of environmental tracers (tritium ^{3}H, helium-3 ^{3}He, chlorofluorocarbons CFC11, CFC12 and CFC113). The flow simulations showed only minor differences in spatial head distribution associated with alternative conceptualizations despite the complexity of the aquifer system and the significant differences in geological interpretations. The models, however, showed major differences in predictions of the age of the groundwater and environmental tracer concentrations, differences that are seen as an effect of model structure uncertainty, because no additional calibrations to these data were performed. A single conceptualization may be adequate in characterizing the natural behavior of a field system after calibration, because the calibration procedure is able to compensate for errors in the data or in the conceptual model through biased parameter values. However, once extrapolation beyond the calibration base is attempted, different conceptual model formulations result in significantly different results. Consequently, it is crucial to take model conceptual uncertainty into account when making predictions beyond the calibration base.

Résumé Quatre modèles conceptuels différents basés sur des interprétations géologiques alternatives ont été élaborés pour un système aquifère phréatique de 600km^2 comprenant des dépôts Quaternaires et situé au Danemark. Chacun des quatre modèles a été calibré par modélisation inverse à partir de niveaux piézomètriques et de mesures de débit. Par la suite, les capacités de transport des quatre modèles ont été comparées à 32 mesures de concentration de traceurs environnementaux (tritium ^{3}H, hélium-3 ^{3}He, chlorofluorocarbures CFC11, CFC12 et CFC113). Malgré la complexité du système aquifère et les différences significatives entre les interprétations géologiques, les simulations d'écoulement n'ont montré que des différences mineures dans la distribution de la piézomètrie en fonction des conceptualisations alternatives. Cependant, les modèles ont montré des différences majeures pour les prédictions de l'âge de l'eau souterraine et pour les concentrations en traceurs environnementaux ; celles-ci sont perçues comme un effet de l'incertitude sur la structure du modèle étant donné qu'aucune calibration supplémentaire n'a été effectuée pour ces données. Après une calibration, une unique conceptualisation peut être adéquate pour caractériser le comportement naturel d'un système de terrain, car la procédure de calibration peut compenser les erreurs liées aux données ou au modèle conceptuel à travers des valeurs de paramètres biaisées. Cependant, une fois l'extrapolation sur la base de la calibration effectuée, des formulations de modèle conceptuel différentes conduisent à des résultats significativement différents. En conséquence, il est crucial de prendre en compte l'incertitude du modèle conceptuel au-delà du processus de calibration pour faire des prédictions.

Resumen Se formularon cuatro modelos conceptuales diferentes, basados en las interpretaciones geológicas alternativas, para un sistema acuífero poco profundo de 600km^2 en Dinamarca, el cual incluye los depósitos Cuaternarios. Cada uno de los cuatro modelos se calibró, contra las mediciones de las cabezas del agua subterránea

Received: 24 June 2005 / Accepted: 25 April 2007
Published online: 26 May 2007

L. Troldborg (✉) · J. C. Refsgaard
Geological Survey of Denmark and Greenland,
Øster Voldgade 10, DK-1350,
Copenhagen, Denmark
e-mail: ltr@geus.dk
Tel.: +45-3814-2000
Fax: +45-3814-2050

J. Refsgaard
e-mail: jcr@geus.dk

K. H. Jensen · P. Engesgaard
Geological Institute,
University of Copenhagen,
Øster Voldgade 10, DK-1350,
Copenhagen, Denmark

K. Jensen
e-mail: khj@geol.ku.dk

P. Engesgaard
e-mail: pe@geol.ku.dk

 DOI 10.1007/s10040-007-0192-y

y del caudal de descarga, a través de modelación inversa. Posteriormente, se compararon las capacidades de transporte de los cuatro modelos con 32 mediciones de concentración de trazadores ambientales (tritio ^{3}H, helio-3 ^{3}He, clorofluorocarbonos CFC11, CFC12 y CFC113). Las simulaciones de flujo mostraron sólo diferencias menores en distribución espacial de cabezas, asociadas con las conceptualizaciones alternativas, a pesar de la complejidad del sistema acuífero y de las diferencias significantes en las interpretaciones geológicas. Los modelos, sin embargo, mostraron las diferencias mayores en las predicciones de la edad del agua subterránea y en las concentraciones del trazador ambiental, diferencias que son vistas como un efecto de incertidumbre de la estructura del modelo, porque no se realizó ninguna calibración adicional a estos datos. Una conceptualización simple puede ser adecuada, caracterizando el comportamiento natural de un sistema de campo después de la calibración, porque el procedimiento de la calibración puede compensar los errores en los datos o en el modelo conceptual a través de valores parcializados del parámetro. Sin embargo, una vez se intenta la extrapolación más allá de la base de la calibración, las formulaciones de modelos conceptuales diferentes producen resultados significativamente diferentes. Por consiguiente, es crucial tener en cuenta la incertidumbre conceptual del modelo, al hacer las predicciones más allá de la base de la calibración.

Keywords Uncertainty · Heterogeneity · Conceptual models · Environmental tracers

Introduction

Groundwater models are increasingly used as support tools for the prediction of concentrations of contaminants. In this context, it is necessary to assess the uncertainties related to model predictions in order for water-resources managers to adequately balance environmental risks and costs and make rational decisions. Predictive uncertainty may originate from three main sources: (1) scenario uncertainty (lack of knowledge on the future evolution of the system), (2) model (conceptual) uncertainty, and (3) parameter uncertainty (Samper et al. 1990). The subject of parameter uncertainty has been addressed by numerous authors, and methods have been developed for dealing with this problem, notably inverse modeling (Christensen and Cooley 1999; Poeter and Hill 1997) and Monte Carlo techniques. Scenario uncertainty can be investigated by testing alternative scenarios.

Methods for assessing conceptual uncertainty are less developed. Martin and Frind (1998) and Beckers and Frind (2001) found that the three-dimensional conceptualization was a key factor for accurate flow field simulations for a moraine aquifer complex in Canada. Sensitivity of the conceptualization was examined in relation to the extent of discrete zones within an overall framework based on hydrostratigraphic interpretation of well log information. Likewise, Fogg (1986) demonstrated the importance of the spatial extent of the geological structures and their interconnectivity in an investigation of a sandstone aquifer system. It was concluded that although the heterogeneity is critically important for three-dimensional transport, it might not always be easily identified based on hydraulic data alone.

In an attempt to eliminate subjectivity of the conceptualization, Tsai et al. (2003) developed a combinatorial optimization scheme of parameter structure identification and applied it to a simple aquifer system. The scheme identifies zonation patterns similar to transmissivity patterns often used when interpreting pumping tests. Sonnenborg et al. (1996) also used a zonation approach along with four model-structure identification criteria to find a balance between model accuracy and the number of parameters to be estimated. From a philosophical point of view, neither of the two papers succeed in the elimination of subjectivity, as in both papers it was subjectively decided that a zonation approach was sufficient to capture the complexity of the system.

Recognizing that subjectivity is an integral part of conceptualization, Harrar et al. (2003) used two alternative geological realizations as the basis for flow modeling of a sand dominated aquifer system in southern Jutland. The simulations showed similar shapes and sizes of particle-tracking simulated-capture zones. The results, however, indicated that travel times and solute breakthrough could vary depending on the conceptualization. Similar results were found in a study by (Zyvoloski et al. 2003). Different geological conceptualizations were developed for a repository site, and it was found that, regardless of geological conceptualization, there was similarity between simulated flow paths.

There are two opposing views on the role of conceptualization. The most predominant approach, described by Beven (2002) as pragmatic realism, regards conceptualization as a continuous process that gradually leads to a conceptual model that is appropriate for the given application. Bredehoeft (2005) compares conceptualization with paradigm theory and, based on case studies, argues that in many instances construction and calibration of models leads to a reformulation of the first conceptualization. This iterative process eventually leads to a singular "best possible representation of the system", Bredehoeft (2003). The second approach essentially regards the concept of conceptualization as an inherently wrong concept. As a conceptual model is a simplification of the real system, it is more relevant to address the span of uncertainty due to an erroneous conceptualization instead of working towards a singular best representation. Carrera et al. (1993) states that "accurate characterization of geological media sounds absurdly utopic", because a conceptual model is based on assumptions on geological variability and patterns, as derived from an often sparse and erroneous database. Consequently, a conceptualization is just one out of several possible alternatives. Likewise, Neuman and Wierenga (2003) and Beven (2002) point out the necessity of analyzing alternative conceptual models in order to explore uncertainty related to conceptualiza-

 DOI 10.1007/s10040-007-0192-y

tion. Although Bredehoeft (2005) finds this second approach appealing, he points out that it has never been "put into practice", and that conceptual models are rarely changed unless challenged by new data.

The objective of this study is to investigate uncertainty related to model conceptualization by evaluating four alternative conceptualizations. The performances of different models, which are based on different conceptualizations and calibrated with respect to head and flow, are explored. Specifically, their predictive capabilities with respect to the simulation of groundwater age and environmental tracer concentrations are evaluated.

Methodology

Four three-dimensional hydrogeological models are compared that only differ on account of the alternative conceptual models of a Quaternary aquifer complex, which is described briefly below. The construction of the four alternative conceptual models is based on different geological interpretations of the aquifer system. Each of the four conceptual models is then introduced in a numerical code and serves as the basis for hydrological modeling.

The modeling procedure is as follows: (1) auto-calibration of a steady-state numerical-groundwater model against observed piezometric heads and stream discharges, (2) simulation of groundwater age distributions at wells using a particle-tracking approach, (3) simulation of environmental tracer concentrations (tritium ^3H, helium-3 ^3He, chlorofluorocarbons CFC-11, CFC-12 and CFC-113) within the groundwater at selected well screen locations using a "convolution" approach by which the simulated age distributions and the temporal variation of tracer concentrations in the water recharging the aquifer system are convoluted. The approach is similar to a blackbox approach, except that in the blackbox approach, the age distribution is estimated a priori, whereas in the "convolution" approach, the age distribution is estimated from the particle-tracking simulation.

A quantitative comparison of the different conceptual models is performed in a two-step procedure. Firstly, the differences in calibrated parameters and residuals among the alternative conceptualizations are investigated. Secondly, each calibrated model is subject to forward particle tracking simulations, and transport times are subsequently compared to age estimates and the concentration measurements of environmental tracers.

Study catchment

Hydrogeology

The study area of Eggeslevmagle (approximately 20× 30 km^2) is located in the western part of the Danish island of Zealand, Fig. 1a. The capture zone for a pumping well field (referred to as the GRUMO area) is located in the central part of the study area, Fig. 1b. Within this area, which is part of the national groundwater monitoring network (GRUMO), most wells are instrumented at several screen depths and frequently monitored. The GRUMO area includes six wells that have been sampled for environmental tracers during the last two decades as part of the monitoring program. Additional monitoring was carried out as part of the present investigation. The locations and well numbers are shown in Fig. 1c.

The landscape is characterized by an undulating topography and a general gradient from the northeastern part of the catchment towards the sea with elevations ranging from 80 to 0 m above sea level (m.a.s.l.) over a distance of 15 km. Vaarby Stream and its numerous tributaries cut through the landscape. Most of the area is artificially drained, but the depth and location of the tile drain system is not known in detail. The tributaries are small and well connected to the drainage system and they respond quickly to rainfall events. The topsoil consists of a clayey and sandy till with a relatively thin unsaturated zone.

Glacial activity during the Main Weichselian (northeast advance) and Young Baltic (southeast advance) and subsequent recessions formed the aquifer system (Hansen 1950; Houmark-Nielsen 1987). The main aquifers consist of sequences of glaciofluvial sand and gravel at varying depths, separated by clayey and sandy glacial tills. The total overburden constitutes 30–80 m of glacial deposits. Tertiary deposits of thick Paleocene clay and Danien limestone are found beneath the Quaternary deposits. The thick Paleocene clay or marl is impervious for all practical purposes. The exchange of water between the Quaternary and the Tertiary deposits was found to be of minor importance to the overall water balance for this part of Zealand (Christensen 1994). The Paleocene clay is lacking in certain areas in the southern part of the system and Danien limestone is found immediately beneath the Quaternary deposits. The aquifer system is important for drinking water supply. Currently, 54 pumping wells are active within the area, many producing more than 50,000 m^3/year, adding up to a total of 4.2 million m^3/year. Most water-producing wells are screened in the Quaternary sandy deposits, but in the southern part, a few wells are screened in the Danien limestone.

The aquifer complex is mainly confined or semi-confined. The typical annual groundwater level fluctuations are 2–3 m. Precipitation and potential evapotranspiration data for 1991–2000 show average annual values of 620 and 530 mm, respectively. For the same period, the average net recharge (precipitation minus actual evapotranspiration) is calculated as 250 mm/year, varying between 160 mm/year (1996) to 330 mm/year (1994; Henriksen et al. 1998).

Previous investigations have shown that the upper till has a large number of fractures and biopores and a relatively high hydraulic conductivity (10^{-4}–10^{-7} m/s) in the 2–8-m-deep oxidized zone. A thin non-continuous heterogeneous layer of gravel and sand is often found embedded in the upper till (Gravesen et al. 2000). This layer, primarily found in the northeastern part of the area,

 DOI 10.1007/s10040-007-0192-y

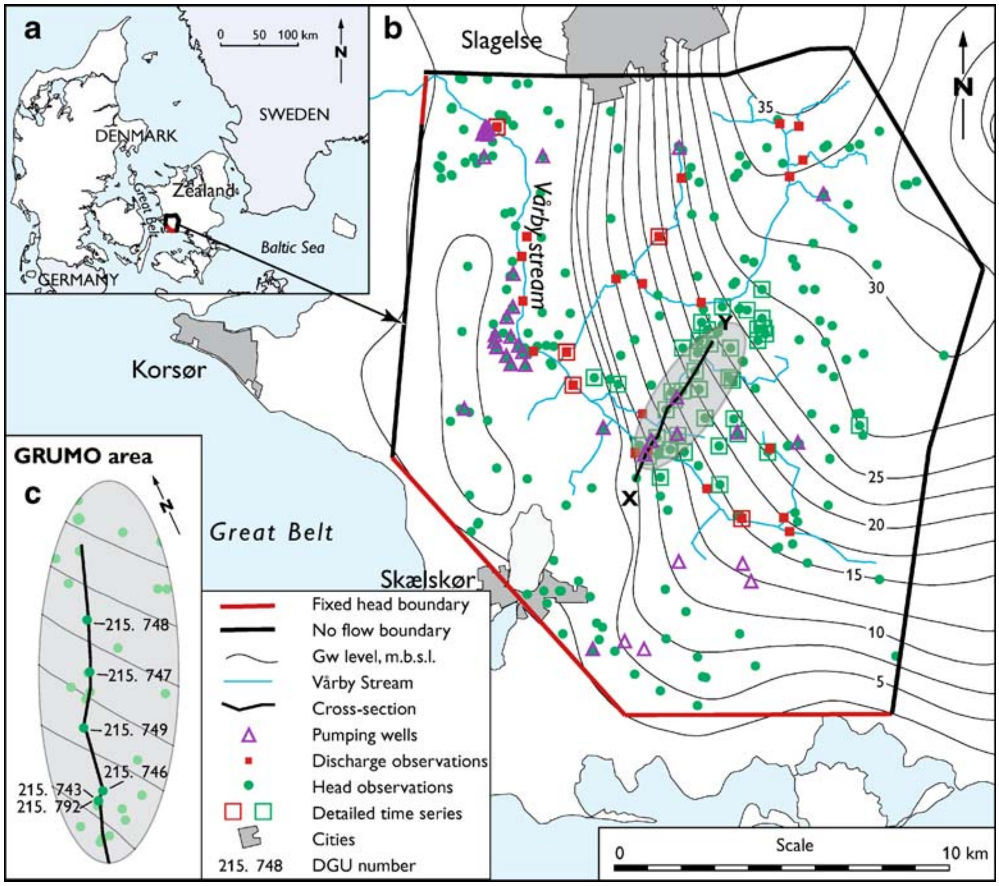

Fig. 1 **a** Location of the study area. **b** Model domain and boundary conditions. **c** The GRUMO (official Danish abbreviation for national groundwater monitoring) area represents the capture zone for a well field; it is part of the national groundwater monitoring network and it is the area from which wells have been sampled for environmental tracers. The *DGU number* refers to the Danish standard for naming wells

often supplies old hand dug wells, but is of minor regional importance, and is therefore not considered further. The till below has not been investigated in detail, but the degree of fracturing is expected to be small (Ernstsen et al. 2000; Klint and Gravesen 1999). Pumping tests conducted in an area of similar geological setting located nearby show an overall vertical hydraulic conductivity for the till in the range 10^{-8}–10^{-9} m/s (Christensen 1994) and horizontal hydraulic conductivities for the glaciofluvial sand and gravel units within the range of 10^{-2}–10^{-5} m/s (Christensen 1997).

Piezometric head and stream discharge data

The piezometric head measurements are divided into two groups: (1) time-series of piezometric head from 58 screens and (2) single or scattered measurements of piezometric head retrieved from the well log archive at the Geological Survey of Denmark and Greenland (GEUS; 256 screens). The first group of data represents wells with detailed well logs and well known vertical positioning of the screens. The second group of data is characterized by (1) a location within or in close proximity to production wells, (2) a single or a limited number of observations that often date back to the time of well construction, (3) uncertainty about screen location, or (4) improper well construction with the possibility for short-circuiting between multiple screens. Evidently, data belonging to this second group are subject to considerable uncertainty. The auto-calibration procedure takes this into account by assigning weight to piezometric heads that are inversely proportional to the uncertainty (variance).

Likewise, stream discharge measurements fall into two groups. One group represents median values of annual daily minimum discharges at 19 different locations, estimated from combining a 20-year series of daily data at a few permanent stations and field campaign data of synchronously measured discharges at the 19 sites in low-flow situations. The second group consists of time series of daily discharge (derived from recordings of water levels that are converted to discharges on the basis of rating curves) at five locations within the same stream system. Both daily values of stream discharge and yearly low-flow measurements are used as calibration targets. As for most Danish streams, overland flow is negligible. Within the study area, the stream is fed by groundwater either directly through the river valley bottom or indirectly through artificial drainage systems.

For model calibration purposes, low-flow measurements are taken as a measure of the baseflow contribution, whereas the daily discharge values are averaged and compared to steady-state simulations of the sum of the aquifer and drainage flow contributions. The uncertainty of the low-flow observations is assumed proportional to the flow value, while the uncertainty of the average flow is taken as the standard deviation of the daily flow for the

DOI 10.1007/s10040-007-0192-y

investigated period (1991–2000). The observed low flows from the 19 locations are in the interval 0–39 l/s and the associated uncertainties are in the range 5–44 l/s. The average flows from the five stations are in the interval 106–1917 l/s with standard deviations of 165–2217 l/s. The locations of the piezometric head and discharge stations are shown in Fig. 1b.

Environmental tracer data

Within the GRUMO area, six wells with 14 screens in total have been sampled for CFC-11, CFC-12, CFC-113, ^{3}H and ^{3}He during four sampling campaigns in the period 1990–2001. The locations of the wells are shown in Fig. 1c. These wells are approximately positioned along a flow line. The water samples for CFC analysis were taken using 45-ml-copper tubes and sealed with pinch-off clamps and subsequently analyzed at the Geological Survey of Denmark and Greenland (GEUS). Helium dissolved gas samples were taken using diffusion samplers and analyzed at Utah University, USA http://www.inscc.utah.edu/~ksolomon/), see Table 1.

The concentrations of environmental tracers have been used for estimating mean ages for groundwater from each of the screens based on a blackbox approach with piston flow assumption (Cook et al. 1995; Table 1). The indicated age estimates are inferred from CFC-12 or tritium (or tritium/helium where available) concentrations. Measurements were also carried out for CFC-11 and CFC-113, but as these tracers are most likely affected by degradation, they were not used for age estimations.

The data suggest that a range of groundwater age estimates can be obtained for the same screens. This is obviously not possible. The vertical positions of the screens together with spans in apparent ages of observed environmental tracers are shown in Fig. 2. The age spans are partly related to difficulties in converting concentrations to ages using simple blackbox methods and partly due to the fact that ages are derived from measurements of different tracers. For a discussion on the temporal instability in age dating using environmental tracers and simple blackbox methods, refer to Troldborg (2004). Although, there are variations from well to well, for example well 215.748 appears to have older water on top of younger water, most of the wells seems to have an age gradient with younger water in the upper screens and older water in the deeper screens. In this part of the area, the upper sandy aquifer is fairly shallow and the observed mean ages may be an indication of a well-mixed aquifer.

Table 1 Measured concentrations of environmental tracers and age estimates based on piston flow assumption

DGU No.	Screen No.	^{3}H	^{3}He	CFC-11	CFC-12	CFC-113	Sample year	Age estimate	Measured by
215.743	3			13.80	4.40	0.00	1998	49	GEUS
		7.81	−6.21				2001	<29	Utah
				103.58	42.90	3.71	2001	35	GEUS
	4	29.00					1990	20	GEUS
				22.40	10.10	11.30	1998	45	GEUS
			5.89				2001		Utah
	5	30.00					1992	25	GEUS
				19.70	13.10	8.20	1998	43	GEUS
		10.71	5.18				2001	7	Utah
				464.72	233.03	44.36	2001	24	GEUS
215.746	1	20.00					1990	<19	GEUS
				1.60	2.10	0.00	1998	51	GEUS
	2			5.50	4.90	0.00	1998	49	GEUS
	3	28.00					1990	19	GEUS
				1638.50	25.50	5.40	1998	38	GEUS
215.747	1	2.60					1990	>40	GEUS
				1.10	0.30	0.00	1998	57	GEUS
		0.62	14.97				2001	58	Utah
				22.91	3.54	0.41	2001	44	GEUS
	2	12.70					1993	<18	GEUS
				0.00	46.00	0.00	1998	34	GEUS
		1.67	24.69				2001	49	Utah
				15.80	3.47	0.00	2001	45	GEUS
215.748	1	0.10	3.45				2001	64	Utah
				55.29	12.15	2.03	2001	39	GEUS
	2	1.00					1993	>40	GEUS
215.749	1			4.50	21.50	0.00	1998	39	GEUS
	2	34.00					1990	37	GEUS
	3	22.00					1992	21	GEUS
				380.90	240.10	24.30	1998	19	GEUS
215.792	1	17.70					1990	<19	GEUS
				0.00	0.30	0.00	1998	57	GEUS

Concentrations for ^{3}H and ^{3}He are in tritium units (TU) and for CFCs in picograms/kg (pg/kg). Age estimates (years) are only based on CFC-12 or ^{3}H (or ^{3}H/^{3}He when available) concentrations. The included DGU number refers to the Danish standard for naming wells. Screens are numbered from the bottom up. Utah refers to Department of Geology and Geophysics at the University of Utah, while GEUS refers to Geological Survey of Denmark and Greenland

 DOI 10.1007/s10040-007-0192-y

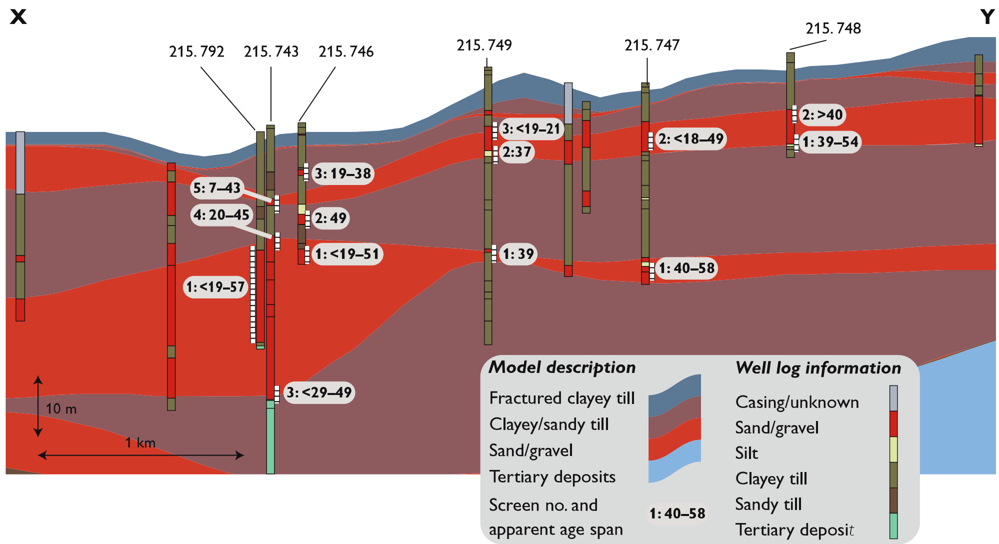

Fig. 2 Geological cross section with well log information and span of apparent groundwater ages (years) derived from tritium and CFC measurements

The large spans in mean age suggest that the results may be affected by errors such as cross-contamination of screens, sampling/analysis errors, or inaccuracies in the applied atmospheric input functions. Poor sampling techniques or cross-contamination of screens due to leaky wells are known to adversely affect the quality of multi-level aquifer concentration data. Five out of 14 screens were pressure tested in an investigation of the monitoring system of the GRUMO area (GEUS 2002). Three (all in well 215.743) of the five screens did not pass the test. However, the leakages from the screens in well 215.743 were so small, that the well was considered suitable for use in the monitoring network. Nevertheless, a minor pollution with CFC or tritium would always lead to the groundwater age being interpreted as later than 1940. The degree to which older wells are affected is not known. Leaks along the well casing would produce higher concentrations of CFCs in the deeper parts, because of increasing atmospheric concentrations. For tritium, the effect could go either way.

Construction of alternative conceptual models

Four different conceptual models based on alternative geological interpretations have been constructed for the study catchment. An overview of the model area is shown in Fig. 1b, while cross-sectional representations of the four alternative geological conceptual models are shown in Fig. 3. Within the study area, the well log archive includes approximately 1,000 wells with detailed geological descriptions and a mean depth of 36 m. The different conceptualizations are primarily based on the well log archive, combined with an interpretation of the depositional processes. They share the same surface topography, surface water bodies, drainage system, recharge, and pumping wells.

The upper till is divided into two sections; an upper highly fractured segment and a lower less fractured segment. The remaining part of the aquifer system consists of units of sand/gravel and glacial clayey or sandy tills. The Tertiary deposit forms the bottom of the active aquifer system and is therefore not included in the conceptualizations.

Model A is a simple aquifer-aquitard model. The depth of the covering layer of fractured and less fractured till is interpolated from well logs. The hydraulic conductivities of the two sections of the till layer are assumed constant and smaller sand/gravel units within the layer are neglected. The sequences of sand, gravel and moraine till below this layer are all treated as a single and uniform aquifer unit with a constant thickness of 25 m. This conceptualization approach is similar to traditional two-layer groundwater models that previously have been successfully used in connection with regional groundwater flow modeling studies in a neighboring area (Christensen 1994).

A variation of this approach is represented by model B, where the same aquifer unit is divided into six zones of different transmissivity. The zonation is based on an interpretation of the depositional environment and geomorphological maps and was performed prior to calibration and not changed during calibration. This is different from the traditional approach, where a feedback mechanism gradually aids in both the forming and shaping of zones. This feedback mechanism is avoided as one of the objectives is to investigate the direct effect of conceptualization.

Model C is a multi-layered aquifer constructed on the basis of well log data that have been aggregated to represent large-scale trends in agreement with the depositional history. The method of aggregation of well log data has been used before in the National Water Resources Model and involves construction of representative wells that integrate well log information for a given area (Henriksen et al. 1998, 2003). The vertical positions of four aquifer units and different clayey units have been interpreted for each of the representative wells. The

 DOI 10.1007/s10040-007-0192-y

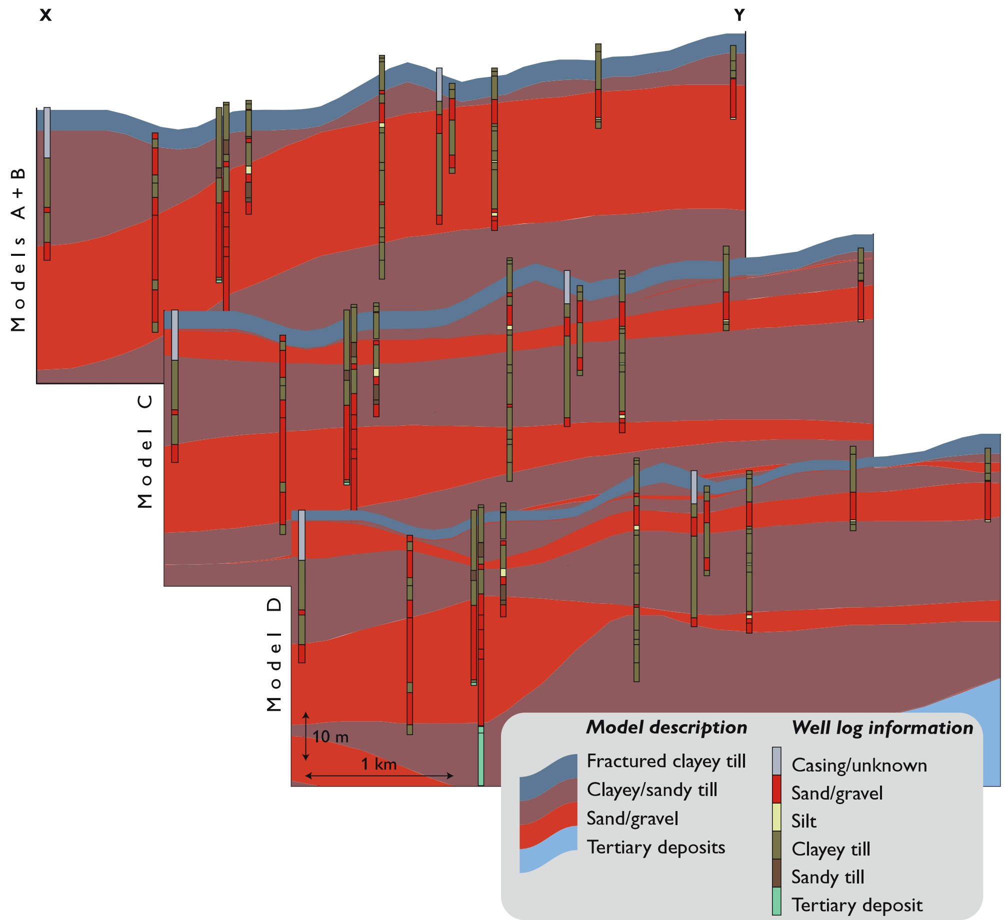

Fig. 3 Cross sections of different hydrogeological conceptualizations. *Model A* is an aquitard-aquifer conceptualization with an aquifer of constant thickness; *model B* is similar to A but the aquifer is divided into six zones; *model C* represents a multi-aquifer system composed of continuous hydrogeological layers that approximate the overall large-scale trends; *model D* is similar to C but honors local well logs

hydrogeology is therefore conceptualized as a continuous layered system consisting of three sand/gravel aquifers embedded in clayey or sandy tills. The continuous sand/gravel aquifers are believed to belong to different glacial advances. The upper till is as in models A and B. However, a thin high-conductivity layer of sand/gravel is embedded in the lower part of the less-fractured till. Some parts of the model area are characterized by a large degree of glaciotectonic variability (Jakobsen 1996) such as push-moraine systems found in the northern part of the area. In these areas, zones of different conductivity break the continuity of the regionally interpreted layered system in the conceptualization. Small-scale variability in sand/gravel and clayey till are not accounted for in the conceptualization and hence layer topography is relative smooth compared to surface topography.

Model D is similar to model C, but interpolation is based directly on well logs rather than on representative wells. The thickness and vertical positions of the different units are thus taken directly from each of the well logs. The upper till is as in the other models, except for the vertical extent of the upper fractured segment, which is now interpreted from local studies of geomorphology, redox conditions and fracture intensity (Klint and Gravesen 1999). A layer of sand/gravel is embedded in the lower part of the upper till as in model C and the remaining part of the system consists of three aquifer units embedded in glacial clayey or sandy tills. Considerable hydrogeological insight is necessary to assure coherency between representative geological cross sections and three-dimensional model conceptualization. In model D, the possibility of "short cuts" between aquifer units exist as they can be in close contact.

Modelling

Setup and calibration of steady-state flow models

The models were constructed using the MIKE SHE modeling system (Abbott et al. 1986). MIKE SHE is a deterministic, fully distributed and integrated hydrological model code capable of simulating the flow of water in the various components of the hydrological cycle including surface water, evapotranspiration, soil water and groundwater (Refsgaard and Storm 1995). As the focus in this study was the impact of the geological conceptualization on the flow pattern and associated travel time distributions within a steady-state flow framework, a number of model simplifications were introduced. The routing of flow in the streams was not considered but streams were included as internal boundaries by specifying water levels along the individual sections. The water exchange between streams and groundwater was simulated based on the prevailing head difference between the two water bodies, stream

DOI 10.1007/s10040-007-0192-y

cross-sectional area and a leakage coefficient. In addition, the diversion of groundwater captured by subsurface drainage system installations to the streams was simulated based on specified drainage depth (m) and drainage constant (s^{-1}). In this manner, the mean discharge along the stream sections was predicted. Flow through the unsaturated zone was not considered and instead a distributed mean groundwater recharge was specified as the upper boundary condition to the groundwater model component. The recharge values were predicted using a stand-alone model that is based on a simple water balance for the root zone where recharge is simulated as the amount of precipitation in excess of actual evapotranspiration and storage in the root zone (Henriksen et al. 2003).

The adopted model domain is shown in Fig. 1b together with the boundary conditions. The boundary conditions are specified as no flow along the topographical divide of the catchment area of Vaarby stream and as fixed heads near the coastline and at the stream valley in the northwestern corner of the model. The bottom of the model follows the top of the Tertiary deposit. The horizontal model discretization is 250 m×250 m, which results in a total of 76×84 cells for each computational layer. These layers follow the hydrogeological units defined for the different conceptualizations and consequently the layering structure will vary among the models. For the particle tracking simulations, the horizontal grid was refined to 125 m (see the following).

The leakage parameter characterizing the exchange of water between groundwater and streams is assumed to be uniform in space and is not subject to calibration. The tile drain system is assumed to be uniformly distributed over the entire area and located 1 m below ground surface and all drainage is directed to the nearest stream location. Fifty-four pumping wells or well fields are included in the model representing more than 99% of the total abstraction within the area and only minor wells used for agricultural purposes are excluded.

The hydrogeological units (fractured clayey till, non-fractured clayey/sandy till, and sand/gravel) in models A, C, and D are given uniform hydraulic conductivity and a constant anisotropy factor ($K_h=10K_v$, where K_h and K_v are the horizontal and vertical hydraulic conductivities, respectively). Model B follows a traditional zonation approach and can be seen as a refinement of model A. The main sand/gravel aquifer is divided into six zones, each with its own hydraulic conductivity, all subject to calibration. The size and location of the zones follow the geomorphologic interpretation. As the focus of the study is on the effect of the conceptualization of the geological structures, porosity variations will not be considered in the simulations and all hydrogeological units are assumed to have a uniform porosity of 0.2.

Each of the individual models based on the alternative geological conceptualizations A–D were auto-calibrated by minimizing an objective function. The objective function was the same for all four models and the weights and observations were identical. The maximum likelihood objective function $S(b)$ specifies the weighted least-square fit of simulated and observed heads and river discharges, respectively, at different locations as a function of parameters being estimated (b), i.e.,

$$S(\underline{b}) = \sum_{i=1}^{NH} \omega_i \left[h_i - h_i'(\underline{b})\right]^2 + \sum_{j=1}^{ND} \omega_j \left[D_j - D_j'(\underline{b})\right]^2$$

where ω represents the weight for observed head (h) [m] and discharge (D) [m^3/s] of the i^{th} or j^{th} observation respectively. NH is the number of head observations, ND is the number of discharge observations. $h_i'(b)$ is the simulated head, and $D_j'(b)$ is the simulated discharge. The number of free parameters is three in models A, C, and D, and eight in model B.

Using head and discharge data in the same objective function requires consideration of the weighting of the two types of data as this will affect the parameter and parameter uncertainty estimates (Anderman and Hill 1999; Barlebo 2000). In common with current practice, the assigned weights were one over the estimated variance of each observation. The universal inverse modeling program UCODE was used for the inversion (Poeter and Hill 1998).

Simulation of age distributions

The groundwater age distribution was simulated using a forward particle-tracking procedure, where particles are released with recharge and moved through the aquifer system based on the simulated local groundwater velocities. Velocities are approximated linearly within each cell based on the midpoint values. The interpolated velocity scheme is mass conservative for purely advective particle tracking (Pollock 1988; Anderson and Woessner 1992; LaBolle et al. 1996). Dispersion and diffusion are not accounted for. As the scheme is mass conservative and particles are added with recharge, it is possible to recalculate particle travel times for a given volume into a groundwater age distribution. The particle tracking simulation is continued until an approximate steady-state situation is obtained, i.e. the difference between the number of particles entering and leaving the model region is less than 1% (occurs after approximately 1,000 years). All particles within a registration radius of 200 m from the screens were used to construct a well-screen groundwater-age distribution (Troldborg 2004).

For computing the age distribution of water reaching a given observation point, the travel times of all the particles that traverse a registration radius around the observation point are recorded over a 500-year-simulation period. The frequency of recorded travel times is subsequently converted into an age distribution (travel time histogram). The calculated age distribution is a result of the spatial variation of the velocity field within the aquifer caused by the geological heterogeneity and the related heterogeneity of the hydraulic parameters.

 DOI 10.1007/s10040-007-0192-y

Derivation of screen specific environmental tracer breakthrough curves

The input functions of the yearly mean concentration of the selected environmental tracers are adapted from a study of a nearby catchment of similar hydrogeological conditions (Troldborg 2004). Similar to that study, the simulations are constrained to the groundwater zone and the interaction with streams. The upper boundary condition is in the form of groundwater recharge and the input functions for the considered tracers should therefore represent the concentration in the recharge water. The basis for estimating these concentrations are records of atmospheric concentrations. The individual tracers considered in this study, their specific attributes and input function characteristics are described in detail in Troldborg (2004).

Gaseous tracers like CFC-11, CFC-12, CFC-113, sulphur hexafluoride SF_6, krypton-85 ^{85}Kr and ^{3}He are in equilibrium with the air phase of the vadose zone. Depending on solubility, diffusion coefficient and thickness of the vadose zone, the air phase may be in equilibrium with the atmosphere (Cook and Bohlke 2000). As the vadose zone in the Vaarby catchment is relatively thin (<5 m) it seems justifiable to assume that the tracer concentrations are not affected during the passage from the soil surface to the water table.

The age distribution computed at each well and the tracer input function at the water table were used to calculate the tracer concentration at the wells as a function of time following the convolution integral procedure discussed by Weissmann et al. (2002) and Troldborg (2004).

The computed age distributions are affected by the size of the grid cells because the particle registration takes place on a cell basis and not on a particle (point) basis. Therefore, it was necessary to change the 250-m grid into a 125-m grid to refine the predictions and to adequately resolve differences in simulated concentrations at closely located observations wells. The effect of changing the grid size without recalibrating the models was tested for changes in the objective functions and water balances. The objective functions for the models with 125-m-grid spacing were equal to or better than the optimized objective function for the same models with 250-m-grid spacing, and the differences in water balances between the various models were within a few mm/year.

Modeling results

Calibration results and analysis

The calibration statistics are shown in Table 2. The value of the objective function shows a clear improvement, mostly in terms of groundwater heads, as the complexity of the geological interpretation increases from models A/B to models C and D. In Fig. 4, the spatial distribution of residuals is shown. The head residuals for all models appear spatially uniformly distributed except in two rather small areas. The northwestern area, which is characterized by large topographical variability and described as a push-moraine landscape, has higher residuals for all the models. The inaccurate location of screens and a high degree of glaciotectonic variability could possibly affect the residuals. The mid-eastern area, which is characterized as a dead-ice landscape, is affected to a large degree by small-scale topographic variability. Models A and B are mostly affected in the northern part, whereas the layered models C and D are mostly affected in the eastern part.

The stream system is gaining from groundwater throughout the area according to both observations and simulations. The stream discharge residuals do not exhibit a notable variation between the models, neither in a global sense (Table 2) nor in a spatial sense (Fig. 4).

Statistics on residuals between simulated and observed heads are compared with residuals between the models in Table 3. Residuals for all models are calculated as the mean of the absolute differences between observed and simulated values. Residuals between models are calculated as the mean of the absolute differences between simulated values by the models being compared. The model-to-model differences (“model2model”) is the average of the six model residuals. It appears from the table that the residuals between observations and simulations by the individual models are notably higher than the residuals between the models, indicating that the differences among the models are generally smaller than the differences between models and observations. The residuals between models A and B are particularly small suggesting that the geological refinement embodied in model B does not have an impact on the simulated heads.

The estimated parameters and their 95% confidence interval are shown in Fig. 5 for the different models. KML1 represents the hydraulic conductivity of the fractured clayey till and is poorly constrained in all four models. KML2 represents the hydraulic conductivity for the remaining part of the clayey or sandy till, while KS1–KS6 represent the hydraulic conductivities for the sand/gravel materials. For the models with only three parameters (A, C and D) both KML2 and KS1 are well determined, whereas model B with eight parameters has larger 95% confidence intervals for KML2 and KS1–KS6 as well. The estimated parameters for KS1–KS6 are in good agreement with the values reported by Christensen (1994) for a similar aquifer.

Comparison between simulated age distributions and apparent age estimates

Figure 6 shows the simulated mean ages for model D for cross-section X–Y together with the ages of particles located within a band of 200 m from the cross section. The mean iso-age lines are hand-contoured based on the particle age pattern. In certain areas, there is a considerable variation in ages. This variability is primarily caused by local depressions in the topography and/or groundwa-

DOI 10.1007/s10040-007-0192-y

Table 2 Hydraulic head and stream flow statistics from model calibration. Root mean squares (RMS) are calculated from 314 head measurements and 24 stream flow measurements

	Units	Model A	Model B	Model C	Model D
Objective function $S(b)$	-	734	708	581	377
RMS of groundwater heads	*m*	5.75	5.61	5.10	4.00
Weighted RMS of groundwater heads	-	1.53	1.50	1.36	1.09
Weighted RMS of stream flow	-	0.27	0.26	0.24	0.23
Error of variance	-	2.19	2.14	1.74	1.12
Number of observations	-	338	338	338	338
Number of parameters	-	3	8	3	3

Error of variance: $s^2=S(b)/(\text{No. of observations}-\text{No. of parameters})$

ter abstractions. It should, however, be noted that the transect is not perfectly aligned with a flow line, which may partly explain some of the shown variability. Generally, there is a gradient in age from the upper part of the aquifer towards the deeper parts. Below local depressions in the topography, the results suggest that particles with longer transport times are more prevalent here. The upconing of ages notable in the left side of all four cross sections are primarily caused by flow from aquifer to stream.

In Fig. 7, the mean age contours are shown for profiles of all four models. For models A and B, the mean ages in the lower part of the aquifer are in the range 50–>500 years while for models C and D, higher ages in

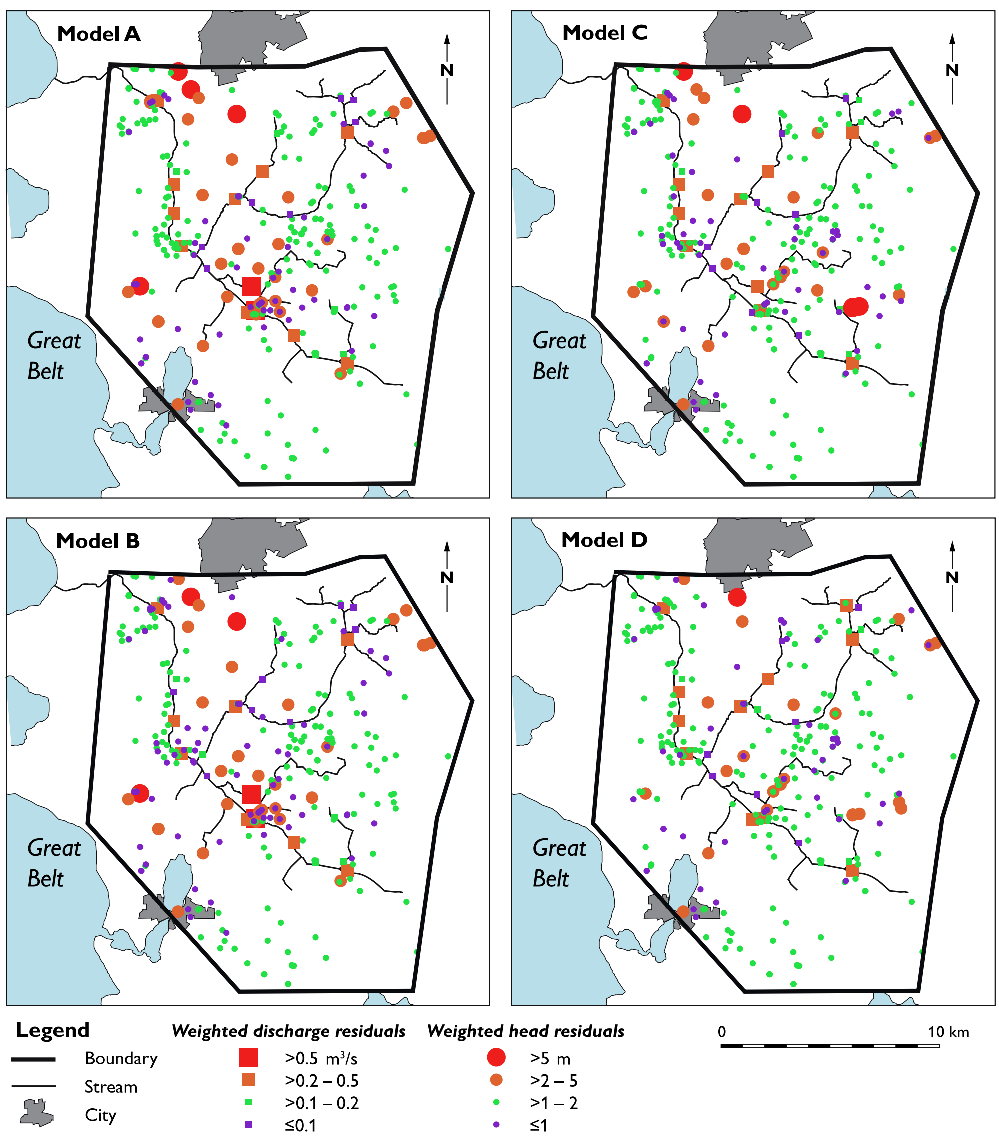

Fig. 4 Spatial variations of head (in meters above mean sea level) and stream discharge residuals for each of the calibrated models

 DOI 10.1007/s10040-007-0192-y

Table 3 Residuals between simulated and observed heads for the four models (*A–D*) and residuals between the individual models (all statistics on absolute values)

	A	B	C	D	AB	AC	AD	BC	BD	CD	Model2 model
Mean residual (m)	3.21	3.32	2.96	2.50	0.54	2.10	2.16	1.96	2.15	1.33	1.71
95 percentile head residual	10.32	10.67	8.85	7.12	2.90	5.35	5.96	5.31	6.02	3.89	5.15
Number of observations	338	338	338	338	338	338	338	338	338	338	2,028

the range of 100–>500 years are simulated. In the upper part of the aquifer system, groundwater ages are less than 25 years for models A, B and C, while for model D, the simulated groundwater ages are more than 25 years. Evidently the more complicated structure embedded in model D, has a notable effect on the simulated groundwater ages.

When comparing observed and simulated mean ages, it appears that incoherent results are obtained. The simple models A and B seem to perform better in simulating the mean age, than models where more geological insight has been used in the conceptualization. The region where models A and B perform better is in the deeper part of the aquifer, where models C and D show deviations of more than 100 years between simulated and observed mean ages. In all simulations, a uniform porosity of 0.2 was used. Since the groundwater ages scale linearly with porosity, younger water could be obtained by decreasing porosity. However, based on the aquifer material and the available geological information, there were no arguments for introducing a smaller porosity.

Comparison between simulated and observed environmental tracer concentrations

The various tracer-input functions (concentrations in groundwater recharge) applied in the simulations for model D at well 215.749 (upper screens 2 and 3) and the predicted breakthrough curves of environmental tracers for models A, B, C and D are shown in Fig. 8, together with the observed concentrations of tritium and CFC-12. Tritium data are from 1990 and 1992, while CFC-12 data is from 1998. The tritium breakthrough curves predicted by the four models A–D all have different shapes which are related to differences in simulated age distributions of the specific screens. It is not possible to evaluate the temporal evolution from just two tritium observations and it is impossible to judge which of the simulation scenarios are closest to reality. It appears, however, that the concentrations simulated by the four models envelop the observations. The CFC-12 breakthrough curves simulated by the four models are almost synchronous in shape, although with different concentration levels, and all having concentrations equal to or smaller than the observed CFC-12 concentration.

Well 215.749 is selected for closer examination for two reasons. Firstly, the four simulated breakthrough curves are selected as typical examples of how the simulated breakthrough curves evolves. Secondly, the simulated breakthrough curves illustrate the value of having measurements of multiple tracers.

Figure 9a, b shows the residuals for all available measurements (time and screen specific) for tritium and CFC-12, respectively. The labels of the *x*-axis link the

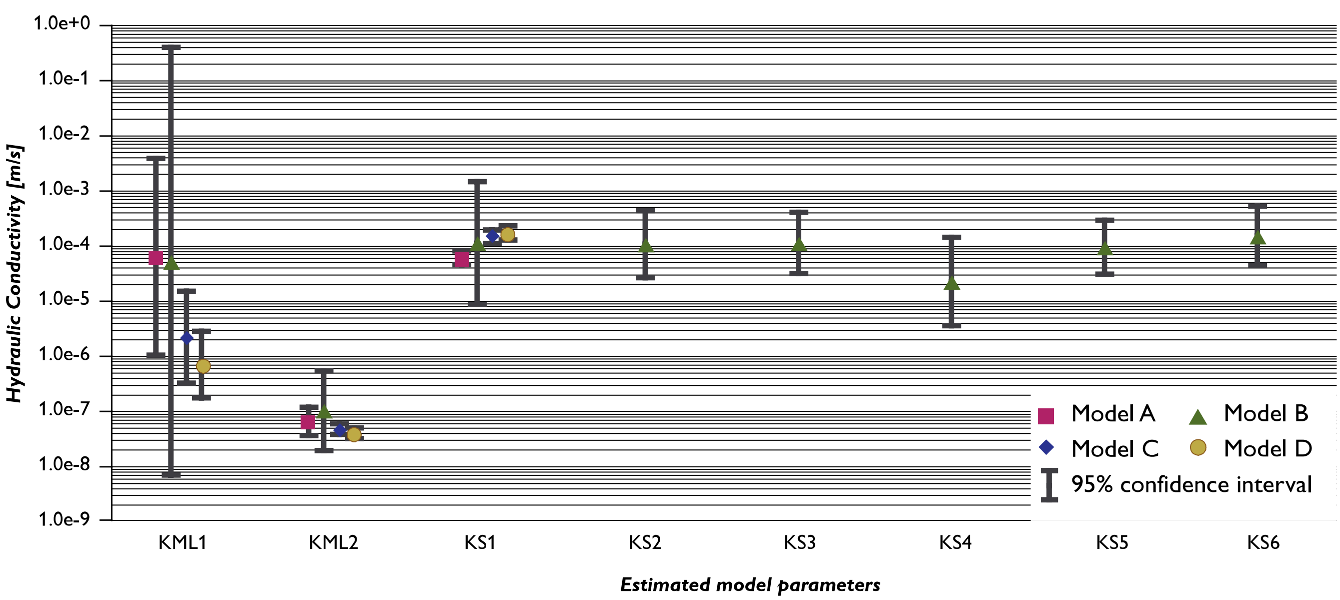

Fig. 5 Estimated parameter values and their 95% confidence intervals

DOI 10.1007/s10040-007-0192-y

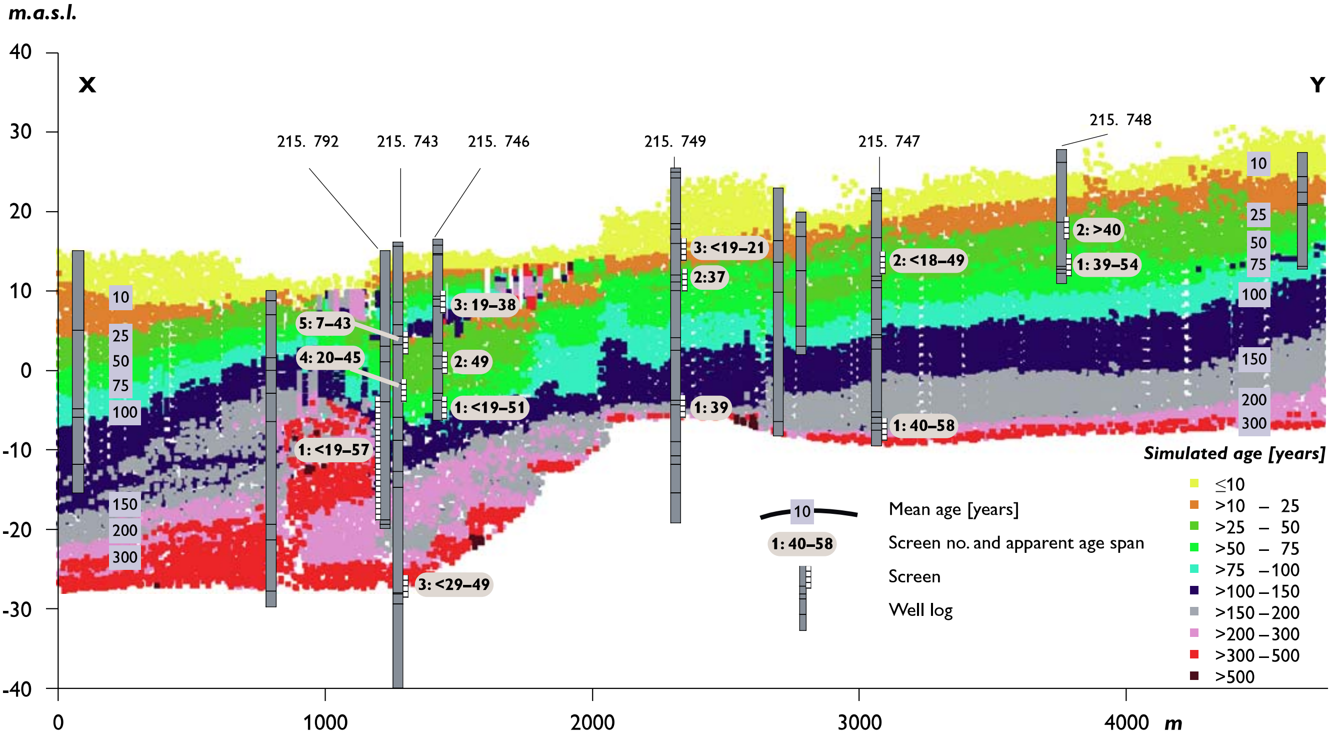

Fig. 6 Cross section (*XY*) with isolines of simulated mean groundwater ages and simulated ages of all particles from within a band of 200m from the cross section (model D) plotted against the observed span of apparent groundwater ages

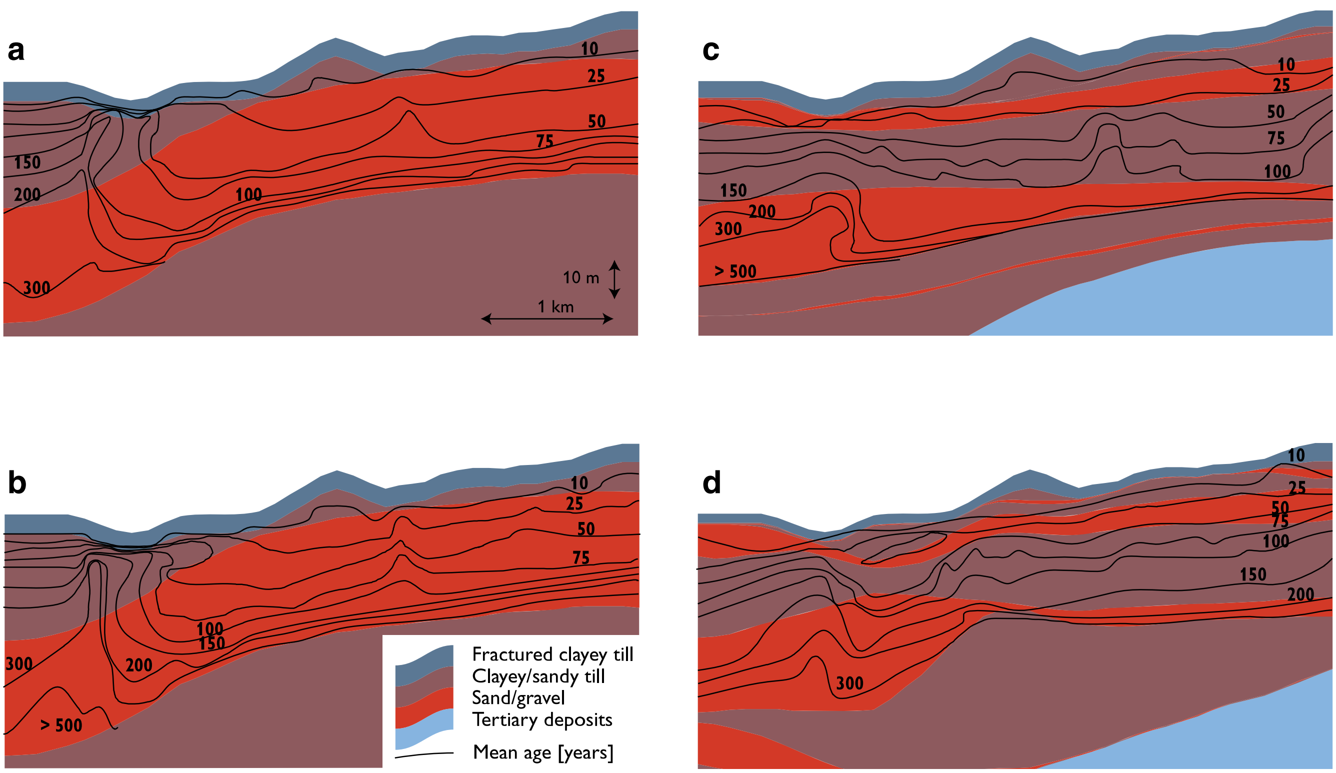

Fig. 7 Cross sections (XY) with isolines of simulated mean ages based on the four alternative conceptual models (*A*, *B*, *C* and *D*). The contour intervals are variable

 DOI 10.1007/s10040-007-0192-y

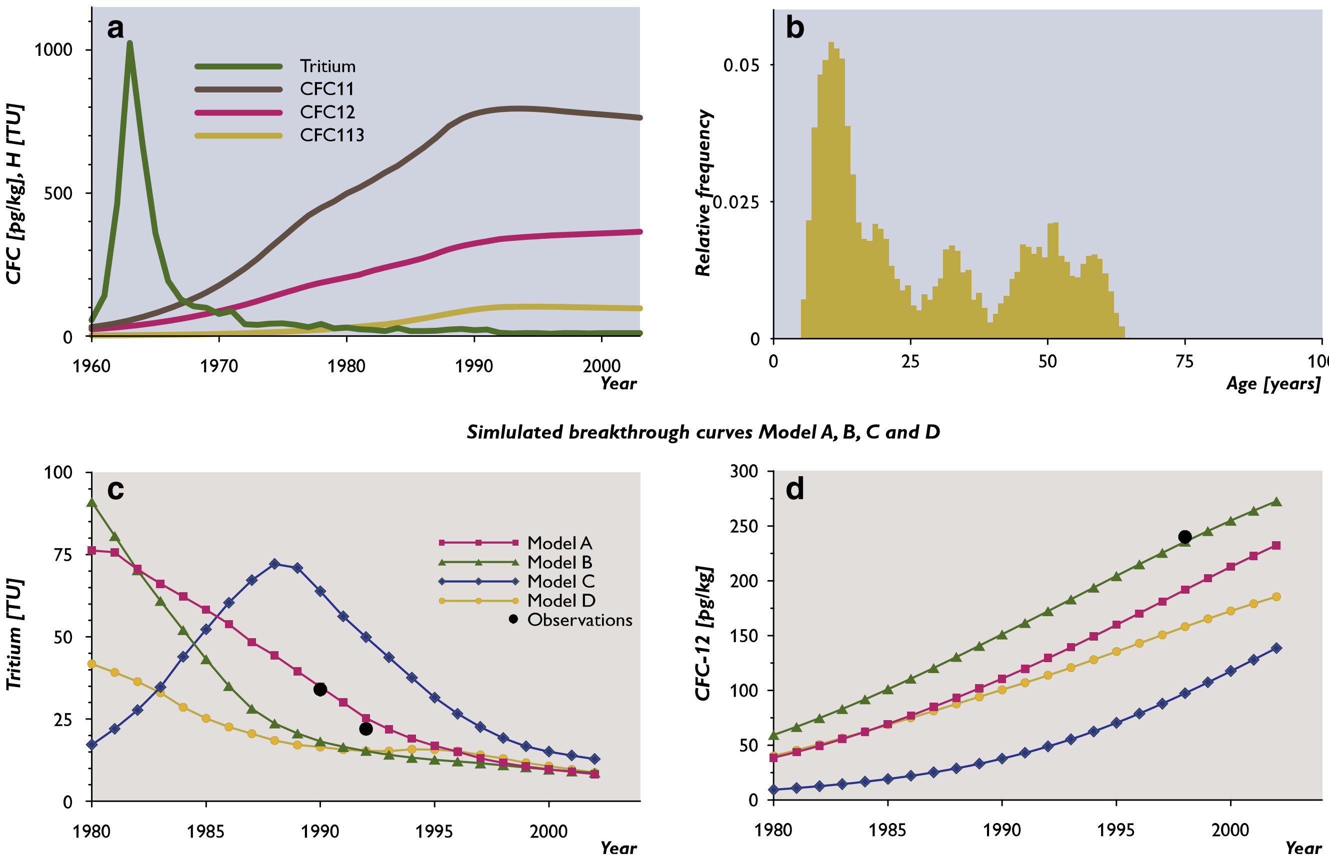

Fig. 8 **a** Input functions of environmental tracers - chlorofluorocarbons (CFCs) in picograms/kilogram (pg/kg) and hydrogen in tritium units (*TU*) for model D at well 215.749, **b** simulated age-distribution at well 215.749 from model *D*, and simulated breakthrough curves for models *A*, *B*, *C* and *D* of **c** tritium and **d** *CFC-12*, together with observed concentrations

residuals to the observation numbers in Table 4. The CFC-11 and CFC-113 results show the same trend as for CFC-12, but these two tracers are to some extent affected by the reduced conditions of the lower parts of the aquifer system and are therefore not shown.

As shown by the figure, the residuals are of the same order of magnitude as the absolute concentrations. Five out of 13 tritium observations and seven out of 13 CFC-12 observations are within or close to the ranges predicted by the four models. Thus, the concentrations simulated by all models encapsulate the observations in about half of the samples. There is no clear trend in the residuals with an almost equal number of positive and negative residuals. However, there seems to be a bias towards positive residuals.

Model D has in most cases smaller residuals than the other models. This is different from the observations of mean age, where models A and B performed better than C and D. Absolute differences between simulated and observed concentrations, as well as between models are presented in Table 5 (“upper” and “lower” refer to screens positioned in the upper and deeper parts of the aquifer, respectively). The lower screens show the smallest differences when comparing model to model. Although comparison between concentration simulations and head and discharge simulations cannot be done directly (e.g. TU compared with m^3/s), the measurements suggest that concentration residuals are one order of magnitude larger than head and discharge residuals, even when measured relative to observation values.

Discussion

Performances of the four models

All the four models perform well with respect to simulation of groundwater heads and discharges in the calibration step. The flow simulations show no major differences in spatial head distribution for the alternative conceptualizations despite the complexity of the aquifer system and the significant differences in geological interpretation. Furthermore, all the calibrated parameter values are within physically realistic ranges. Hence, the underlying four conceptual models can be claimed to be corroborated in the sense that they were not contradicted by field data. A strict quantitative comparison of the four models (Table 2) suggests that model D is marginally better than the other three models. An important characteristic is that the deviations between models are smaller than the deviations between models and observations (Table 3).

 DOI 10.1007/s10040-007-0192-y

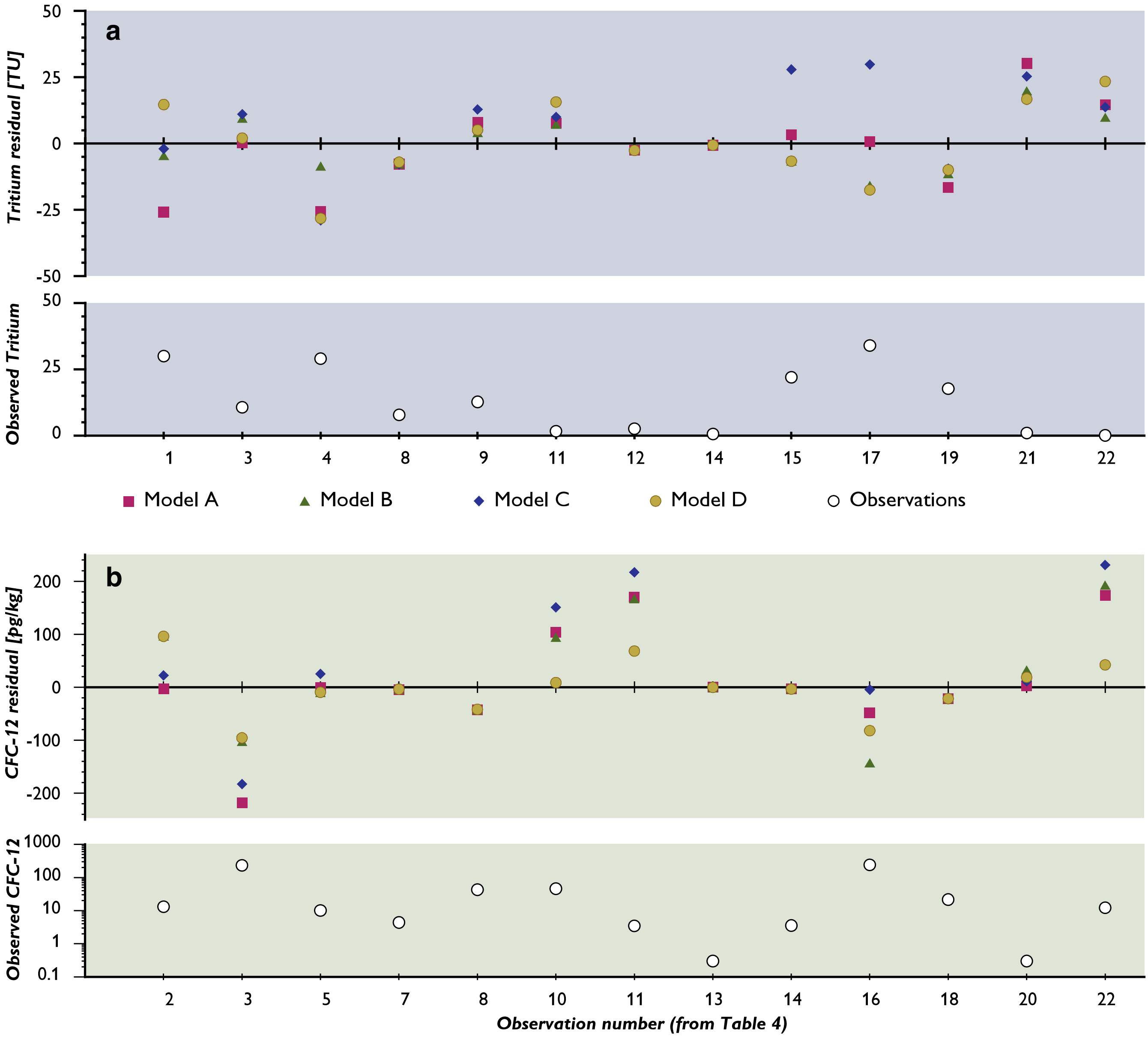

Fig. 9 Residuals between observed and simulated concentrations of **a** tritium and **b** CFC-12 for the alternative models and observed concentrations for the observation numbers listed in Table 4

The flow field is indirectly linked to spatial differences in discharge, especially in the upper part of the aquifer system due to the artificial drainage system. The discharge simulations are, similarly to the flow field, relatively insensitive to changes in conceptualization. In all cases, however, the overall simulated exchange with the groundwater system is within a reasonable range. The similarity in residual distribution between the models (Fig. 4) suggest that other factors, not found important in conceptualization, may influence the flow field.

Models A and B produce younger groundwater ages in the deeper parts of the aquifer system than models C and D. Introducing more layers probably affects transport behavior in two ways. Firstly, it enhances horizontal flow, with only a small percentage of water flowing to the deeper parts of the system, and, in consequence, the simulated transport times to the deeper aquifers are increased. Secondly, in accordance with the theory of Toth (1999) and Zijl (1999), layering dampens the effect of small-scale topographical variation. In simple cases, the topographical effect on the flow field is apparent almost all the way to the deeper parts of the aquifer, whereas it is limited to the upper parts in the more complex cases of models C and D.

The model performance for apparent ages shows better agreement for the simple models (A and B), that at the same time had the poorest head and discharge performance. Previous studies (Weissmann et al. 2002) have

 DOI 10.1007/s10040-007-0192-y

Table 4 Comparison of measured (*Obs.*) and simulated (*A-D*) concentrations. Observation numbers (No.) are included as references for comparison with Fig. 9

				^{3}H (TU)					^{3}He (TU)					CFC-11 (pg/kg)					CFC-12 (pg/kg)					CFC-113 (pg/kg)				
DGU No.	Screen No.	No.	Year	Obs	A	B	C	D	Obs.	A	B	C	D	Obs.	A	B	C	D	Obs.	A	B	C	D	Obs.	A	B	C	D
215.743	5	1	1992	30.0	4.1	25.4	28.0	44.7																				
		2	1998											19.7	13.5	66.0	243.3	244.4	13.1	9.6	109.4	35.3	109.2	8.2	0.7	3.0	21.3	13.9
		3	2001	10.7	10.9	20.2	21.8	12.7	5.2	87.5	135.2	133.8	60.3	464.7	23.1	101.2	291.2	315.4	233.0	14.6	130.5	50.1	137.4	44.4	1.2	4.5	26.1	20.5
	4	4	1990	29.0	3.3	20.5	0.0	0.7																				
		5	1998											22.4	13.5	66.0	0.0	0.8	10.1	9.6	0.0	35.3	0.6	11.3	0.7	3.0	0.0	0.0
		6	2001						5.9	87.5	135.2	10.0	10.4															
	3	7	1998											13.8	0.0	0.0	0.0	0.8	4.4	0.0	0.0	0.0	0.6	0.0	0.0	0.0	0.0	0.0
		8	2001	7.8	0.0	0.0	0.0	0.7						103.6	0.0	0.0	0.0	1.4	42.9	0.0	0.0	0.0	1.0	3.7	0.0	0.0	0.0	0.1
215.747	2	9	1993	12.7	20.7	16.9	25.5	17.8																				
		10	1998											0.0	351.7	465.7	326.6	113.9	46.0	149.9	140.4	196.6	54.6	0.0	29.3	43.6	17.9	5.9
		11	2001	1.7	9.4	8.9	11.7	17.3	24.7	35.3	24.7	48.6	119.0	15.8	406.1	516.7	402.2	155.6	3.5	173.2	170.4	220.4	71.7	0.0	38.0	53.8	26.4	8.8
	1	12	1990	2.6	0.0	0.1	0.0	0.0																				
		13	1998											1.1	0.0	0.0	0.0	0.0	0.3	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0
		14	2001	0.6	0.0	0.1	0.0	0.0	15.0	10.0	9.9	10.0	10.0	22.9	0.0	0.1	0.0	0.0	3.5	0.0	0.0	0.3	0.0	0.4	0.0	0.0	0.0	0.0
215.749	3	15	1992	22.0	25.3	15.4	49.9	15.3																				
		16	1998											380.9	454.6	563.2	212.3	372.9	240.1	191.7	97.4	235.6	158.1	24.3	34.7	50.9	9.2	35.1
	2	17	1990	34.0	34.8	18.2	63.8	16.4																				
	1	18	1998											4.5	0.0	0.1	0.0	0.0	21.5	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0
215.792	1	19	1990	17.7	1.1	6.3	8.2	7.7																				
		20	1998											0.0	4.2	20.6	73.2	41.9	0.3	3.0	32.9	11.4	18.9	0.0	0.2	1.0	6.4	2.4
215.748	2	21	1993	1.0	31.2	20.9	26.3	17.7																				
	1	22	2001	0.1	14.7	10.0	13.8	23.5	3.5	65.3	25.5	59.4	165.3	55.3	430.5	574.7	481.5	109.8	12.1	185.0	205.1	243.1	54.3	2.0	36.7	53.2	44.0	5.7

shown that the simple method of age comparison has some disadvantages because age estimates are sensitive to sampling date and biased in space when applied to complex flow systems, as in this case. Troldborg (2004) argued that the most important error associated with direct age comparison is caused by the underlying assumptions of the flow field usually made when converting environmental tracer concentrations to apparent ages, namely piston flow without dispersion. The shift between the 50–75 isoline year roughly marks the lower limit for simple age estimates using CFCs and tritium/helium. For models A and B, the 50–75 shift lies in the middle of the aquifer (Fig. 7). For model C, the 50–75 shift lies below the upper aquifer, whereas model D shows the 50–75 shift within the upper aquifer. Both C and D show an aquifer with a groundwater age of more than 100 years. It is likely that part of the reason for the better agreement between simulated and apparent ages for models A and B arises from the similarity in assumptions used for calculating apparent age and assumptions regarding the flow field in models A and B. Thus, the simple models represent a flow system that is well in agreement with the piston flow assumption, whereas this is not the case for more complex models like C and D. The underlying assumption of flow conditions associated with groundwater age dating is therefore likely to constrain the usefulness of age predictions from environmental tracers, especially in connection with model calibration/validation.

Some of these problems can be resolved if comparisons are made directly with observed concentrations of environmental tracers. The method used here is essentially equivalent to standard solute transport simulation. In this way, the mismatch of assumptions behind modeling and observations that significantly confuses the interpretation of the groundwater age simulations is avoided. In this case, it turned out that model D, having the worst fit of mean ages but the best of heads and flows, appears to have a better fit than the other models with respect to concentrations, especially with regard to the upper screens.

Comparing differences between the models

The accuracy of the predictions of groundwater ages and concentrations was not impressive for any of the four models. None of the models provided accurate concentration predictions for all screens, but, nevertheless, the simulations did to some extent encapsulate the observation points. The four models were conceptualized according to geological knowledge and with the experience that such conceptualizations have previously given acceptable model predictions of groundwater heads and discharge. The present results suggest that the predictive capabilities of such groundwater transport models are questionable for point-scale concentrations, despite the underlying flow models performing adequately with respect to well heads and stream discharges. Therefore, uncertainty assessments of model predictions are particularly important in case of

 DOI 10.1007/s10040-007-0192-y

Table 5 Mean residuals between simulated and observed tritium and CFC-12 concentrations for the four models and residuals between the individual models (all statistics on absolute values)

		A	B	C	D	AB	AC	AD	BC	BD	CD
Tritium (TU)	Upper	11.9	9.9	15.8	12.4	9.1	12.0	12.3	11.9	6.2	15.5
	Lower	9.2	4.8	10.0	9.6	4.4	0.8	0.8	5.2	5.2	0.4
	All	11.1	8.3	14.0	11.6	7.6	8.6	8.8	9.8	5.9	10.9
CFC-12 (pg/kg)	Upper	90.0	105.6	104.8	52.6	36.4	47.7	76.1	61.7	95.1	52.2
	Lower	14.5	14.4	14.5	14.2	0.1	0.0	0.3	0.1	0.4	0.3
	All	61.0	70.5	70.1	37.8	22.4	29.4	46.9	38.0	58.7	32.2

solute transport modeling, and, preferably, the accuracy of model predictions should be tested against independent field concentration data.

If no concentration data had existed, all four models in this study would be considered as plausible and equally suitable for making predictions, because they all performed well in the calibration step. The differences among the four model simulations can thus be seen as an effect of model structure uncertainty. A relevant question, then, is whether the four alternative conceptualizations under- or overestimates the model structure uncertainty.

Comparison of model-to-model deviations and model-to-observation deviations of groundwater heads in the calibration step (Table 3) indicates a relative closeness between the models. Thus, none of the four models are as far apart from each other than the best model is from the observations. This might indicate that the conceptualizations do not capture all the important features of the hydrogeology and that the differences between the four models are not large enough to adequately represent the diversity within the plausible space of conceptualizations. This indication is confirmed by the qualitative comparisons of the spatial distribution of the head and discharge residuals.

A similar comparison of model-to-model and model-to-observation deviations of transport-results (Table 5) shows that the models are much closer to each other for predictions of environmental tracer concentrations in the lower screens than they are to the observations. Differences between models for predictions of concentrations in the upper screens are similar to the deviations between models and observations. This is in agreement with the above observation that the four hydrogeological interpretations are not sufficiently different to adequately represent the plausible space of conceptualizations.

Conclusion

The four models were calibrated against field data for groundwater heads and discharges, and subsequently tested against independent data for groundwater ages and environmental tracer concentrations. The simulations of ages and concentrations represent an extrapolation step beyond the calibration base. The differences in conceptualizations have been shown to affect the simulations in both the calibration and the extrapolation steps. This provides an opportunity to analyze the importance of model structure uncertainty when making extrapolatory predictions.

First of all, it is noticed that although the simulation results of heads and discharges are similar for the four models, the differences among the four models' simulated ages and concentrations are very significant. This corroborates the findings of Harrar et al. (2003), who investigated the benefits of using alternative conceptual models in capture zone and transport modeling of a relatively simple aquifer system in the western part of Denmark. Their results indicated little difference in model performance with regard to shape and size of simulated capture zones for models calibrated against head and flow observations. However, when using the models for extrapolation beyond the calibration base for prediction of groundwater ages and breakthrough curves, the difference between the models became much greater.

Secondly, the prediction range, in this case ability to reproduce tracer concentrations, produced by the four model simulations, apparently underestimates the effects of model structure uncertainty in the present case. This study clearly shows the effects of calibration. For simulation of heads and discharges, the differences in conceptualizations (and hence the model structure errors resulting in conceptual uncertainty) are compensated for by biases in the parameter values. Thus, the conceptual uncertainty has relatively small effect for predictions within the calibration regime. When the models are used for extrapolations beyond the data basis used for calibration the differences due to different conceptualizations become important. This is in agreement with Harrar et al. (2003) as well as Hojberg and Refsgaard (2005) and Neuman and Wierenga (2003) who recognized conceptual uncertainty as the dominant source of uncertainty in model predictions. It has been shown that the importance of the conceptual uncertainty increases with the range of extrapolation.

Establishment of a conceptual model involves interpretation and integration of different kinds of data including both hard and soft data. The differences in conceptualization are based on subjective hydrogeological interpretations, and subjectivity is recognized as an integrated part of constructing a conceptual model (Fogg 1986; Harrar et al. 2003; Neuman and Wierenga 2003). Because of the subjectivity involved, it is not possible to describe exactly to what extent a sample of alternative

 DOI 10.1007/s10040-007-0192-y

conceptual models is sufficient to explore the plausible space of conceptualizations or which of the conceptual models are best. Evaluation of alternative conceptual models should therefore include both quantitative and qualitative elements.

Acknowledgements We wish to thank K. Solomon and A. Rigby both of the Department of Geology and Geophysics at the University of Utah for important contributions in the form of analysis of tritium and helium and for their insight into sampling of helium by diffusion samplers. Their generous loan of equipment and introduction to sampling associated with diffusion samplers is greatly appreciated. We are grateful to T. Laier for CFC analysis, to P.R. Jakobsen for the geological interpretation of model D and to P. Nyegaard for the geological interpretation of model C as well as guidance in understanding the depositional history of the area. This work was partially funded under a contract with the Ministry of Energy and Environment, 1996–2001, for the establishment of a National Water Resource Model (DK-model) and partially funded by the Danish Research Agency.

References

Abbott MB, Bathurst JC, Cunge JA, Connell PE, Rasmussen J (1986) An introduction to the European Hydrological System-Systeme Hydrologique Europeen, “SHE”: 2, structure of a physically-based, distributed modelling system. J Hydrol 87:61–77
Anderman ER, Hill MC (1999) A new multistage groundwater transport inverse method: presentation, evaluation, and implications. Water Resour Res 35:1053–1063
Anderson MP, Woessner WW (1992) Applied groundwater modeling. Academic, San Diego, CA, USA
Barlebo HC (2000) Parameter identification in groundwater modelling (inverse modelling). PhD Thesis, ISVA series paper no. 75, Technical University of Denmark, Denmark
Beckers J, Frind EO (2001) Simulating groundwater flow and runoff for the Oro Moraine aquifer system. Part II. Automated calibration and mass balance calculations. J Hydrol 243:73–90
Beven K (2002) Towards a coherent philosophy for modelling the environment. Proc R Soc Lond A Math Phys Eng Sci 458 (2026):2465–2484
Bredehoeft JD (2003) From models to performance assessment: the conceptualization problem. Ground Water 41(5):571–577
Bredehoeft J (2005) The conceptualization model problem-surprise. Hydrogeol J 13(1):37–46
Carrera J, Mousavi SF, Usunoff E, Sanchez-Vila X, Galarza G (1993) A discussion on validation of hydrogeological models. Reliab Eng Syst Saf 42:201–216
Christensen S (1994) Hydrological model for the Tude: a catchment. Nord Hydrol 25:145–166
Christensen S (1997) On the strategy of estimating regional-scale transmissivity fields. Ground Water 35:131–139
Christensen S, Cooley RL (1999) Evaluation of prediction intervals for expressing uncertainties in groundwater flow model predictions. Water Resour Res 35:2627–2639
Cook PG, Bohlke JK (2000) Determining timescale for groundwater flow and solute transport. In: Cook PG, Herczeg AL (eds) Environmental tracers in subsurface hydrology: 1–30. Kluwer, Dordrecht, Holland
Cook PG, Solomon DK, Plummer LN, Busenberg E, Schiff SL (1995) Chlorofluorocarbons as tracers of groundwater transport processes in a shallow, silty sand aquifer. Water Resour Res 31:425–434
Ernstsen V, Jacobsen OS, Elsgaard L, Fomsgaard I, Vinther FP (2000) Geokemi, omsætning og sorption af pesticider i den umættede zone [Geochemistry, reduction and sorption of pesticides in the vadose zone]. Miljøforskning 42(1):10–13
Fogg GE (1986) Groundwater-flow and sand body interconnectedness in a thick, multiple-aquifer system. Water Resour Res 22:679–694
GEUS (2002) Grundvandsovervågningsboringers egnethed til analyse [Suitability of groundwater monitoring wells in analysis]. Geological Survey of Denmark and Greenland, GEUS, Copenhagen, Denmark
Gravesen P, Klint KES, Jacobsen OS, Harrar WG, Fleding G, Jørgensen P, Fomsgaard I, Spliid NH (2000) Transport af vand og pesticider i opsprækket moræneler [Transport of water and pesticides in fractured clayey till]. In: Nielsen HØ (ed) Miljøforskning 42(1):6–9
Hansen S (1950) Ekskursion til Slagelse-Korsoer-egnen [Excursion to the Slagelse-Korsoer area]. Bull Geol Soc Den 11:603–607
Harrar WG, Sonnenborg TO, Henriksen HJ (2003) Capture zone, travel time, and solute-transport predictions using inverse modeling and different geological models. Hydrogeol J 11:536–548
Henriksen HJ, Troldborg L, Knudby CJ, Dahl M, Nygaard P, Jakobsen PR, Rasmussen P (1998) National water resources model: Zealand, Lolland, Falster and Mon (in Danish). Geological Survey of Denmark and Greenland Report 1998/109. GEUS, Copenhagen, Denmark
Henriksen HJ, Troldborg L, Nygaard P, Sonnenborg TO, Refsgaard JC, Madsen B (2003) Methodology for construction, calibration and validation of a national hydrological model for Denmark. J Hydrol 280:52–71
Hojberg AL, Refsgaard JC (2005) Model uncertainty: parameter uncertainty versus conceptual models. Water Sci Technol 52 (6):177–186
Houmark-Nielsen M (1987) Pleistocene stratigraphy and glacial history of the central part of Denmark. Bull Geol Soc Den 36:1–189
Jakobsen PR (1996) Distribution and intensity of glaciotectonic deformation in Denmark. Bull Geol Soc Den 42(2):175–185
Klint KES, Gravesen P (1999) Fractures and biopores in Weichselian clayey till aquitards at Flakkebjerg, Denmark. Nord Hydrol 30:267–284
LaBolle EM, Fogg GE, Tompson AFB (1996) Random-walk simulation of transport in heterogeneous porous media: local mass-conservation problem and implementation methods. Water Resour Res 32:583–593
Martin PJ, Frind EO (1998) Modeling a complex multi-aquifer system: the Waterloo Moraine. Ground Water 36:679–690
Neuman SP, Wierenga PJ (2003) A comprehensive strategy of hydrogeologic modeling and uncertainty analysis for nuclear facilities and sites. NUREG/CR-6805, University of Arizona, USA. Available online. http://www.nrc.gov/reading-rm/doc-collections/nuregs/contract/cr6805/. Cited 5 May 2006
Poeter EP, Hill MC (1997) Inverse models: a necessary next step in groundwater modeling. Ground Water 35:250–260
Poeter EP, Hill MC (1998) Documentation of UCODE, a computer code for universal inverse modeling. Water-Resources Investigation Report 98-4080, USGS, Denver, Colorado
Pollock DW (1988) Semianalytical computation of path lines for finite-difference models. Ground Water 26(6):743–750
Refsgaard JC, Storm B (1995) MIKE SHE. In: Singh VP (ed) Computer models of watershed hydrology. Water Resources Publications, Highlands Ranch, CO, USA, pp 809–846
Samper J, Carrera J, Galarza G, Medina A (1990) Application of an automatic calibration technique to modelling an alluvial aquifer. IAHS-AISH Publ no. 195, IAHS, Wallingford, Oxfordshire, UK, pp 87–95
Sonnenborg TO, Engesgaard P, Rosbjerg D (1996) Contaminant transport at a waste residue deposit. 1. Inverse flow and nonreactive transport modeling. Water Resour Res 32:925–938
Toth J (1999) Groundwater as a geologic agent: an overview of the causes, processes, and manifestations. Hydrogeol J 7:1–14
Troldborg L (2004) The influence of conceptual geological models on the simulation of flow and transport in Quaternary aquifer systems, PhD Thesis, Geological Survey of Denmark and Greenland Report 2004/107, GEUS, Copenhagen, Denmark

DOI 10.1007/s10040-007-0192-y

Tsai FTC, Sun NZ, Yeh WWG (2003) A combinatorial optimization scheme for parameter structure identification in ground water modeling. Ground Water 41:156–169

Weissmann GS, Zhang Y, LaBolle EM, Fogg GE (2002) Dispersion of groundwater age in an alluvial aquifer system. Water Resour Res 38(16):1–13

Zijl W (1999) Scale aspects of groundwater flow and transport systems. Hydrogeol J 7:139–150

Zyvoloski G, Kwicklis E, Eddebbarh AA, Arnold B, Faunt C, Robinson BA (2003) The site-scale saturated zone flow model for Yucca Mountain: calibration of different conceptual models and their impact on flow paths. J Contam Hydrol 62–63:731–750

 DOI 10.1007/s10040-007-0192-y