

Assessment of conceptual model uncertainty for the regional aquifer Pampa del Tamarugal – North Chile

R. Rojas^{1,*}, O. Batelaan^{1,2}, L. Feyen³, and A. Dassargues^{1,4}

¹Applied geology and mineralogy, Department of Earth and Environmental Sciences, Katholieke Universiteit Leuven, Celestijnenlaan 200E, 3001 Heverlee, Belgium

²Department of Hydrology and Hydraulic Engineering, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium

³Land management and natural hazards unit, Institute for Environment and Sustainability (IES), Joint Research Centre (JRC), European Commission, Via E. Fermi 2749, TP261, 21027 Ispra (Va), Italy

⁴Hydrogeology and Environmental Geology, Department of Architecture, Geology, Environment, and Constructions (ArGenCo), Université de Liège, B.52/3 Sart-Tilman, 4000 Liège, Belgium

*now at: Land management and natural hazards unit, Institute for Environment and Sustainability (IES), Joint Research Centre (JRC), European Commission (EC), Via E. Fermi 2749, TP261, 21027 Ispra (Va), Italy

Received: 24 August 2009 – Published in Hydrol. Earth Syst. Sci. Discuss.: 17 September 2009

Revised: 7 January 2010 – Accepted: 14 January 2010 – Published: 2 February 2010

Abstract. In this work we assess the uncertainty in modelling the groundwater flow for the Pampa del Tamarugal Aquifer (PTA) – North Chile using a novel and fully integrated multi-model approach aimed at explicitly accounting for uncertainties arising from the definition of alternative conceptual models. The approach integrates the Generalized Likelihood Uncertainty Estimation (GLUE) and Bayesian Model Averaging (BMA) methods. For each member of an ensemble M of potential conceptualizations, model weights used in BMA for multi-model aggregation are obtained from GLUE-based likelihood values. These model weights are based on model performance, thus, reflecting how well a conceptualization reproduces an observed dataset D . GLUE-based cumulative predictive distributions for each member of M are then aggregated obtaining predictive distributions accounting for conceptual model uncertainties. For the PTA we propose an ensemble of eight alternative conceptualizations covering all major features of groundwater flow models independently developed in past studies and including two recharge mechanisms which have been source of debate for several years. Results showed that accounting for heterogeneities in the hydraulic conductivity field (a) reduced the uncertainty in the estimations of parameters and state variables, and (b) increased the corresponding model weights used for multi-model aggregation. This was more noticeable when the hydraulic conductivity field was conditioned on available hydraulic conductivity measurements. Contri-

bution of conceptual model uncertainty to the predictive uncertainty varied between 6% and 64% for ground water head estimations and between 16% and 79% for ground water flow estimations. These results clearly illustrate the relevance of conceptual model uncertainty.

1 Introduction and scope

Due to a lack of precipitations and high evaporation rates, groundwaters in Northern Chile are a strategic source of freshwater. One of the most important groundwater reserves in this region is the regional aquifer contained in the recent deposits of the Pampa del Tamarugal basin (Fig. 1). In this basin, annual precipitation is nil at altitudes below 2000 m a.s.l. but reaches values of about 200 mm yr⁻¹ at altitudes above 3500 m a.s.l. This spatial pattern of the precipitation together with high evaporation rates controls the hydrology of the region. Groundwater and surface water originating in the Andes Mountains are the main water sources for human activities (Aravena, 1995).

In this region, many local towns as well as the capital city of the province (Iquique), most of the mining industry, and a large part of the agricultural sector, entirely depend on groundwater resources. As a result, groundwater abstraction rates from the Pampa del Tamarugal Aquifer (PTA) have gradually increased since the early 1960s resulting in a steady decrease in the groundwater heads recorded in the monitoring network controlled by the Dirección General de Aguas de Chile (DGA) (e.g. Rojas and Dassargues, 2007).



Correspondence to: R. Rojas
(rodrigo.rojas@jrc.ec.europa.eu)

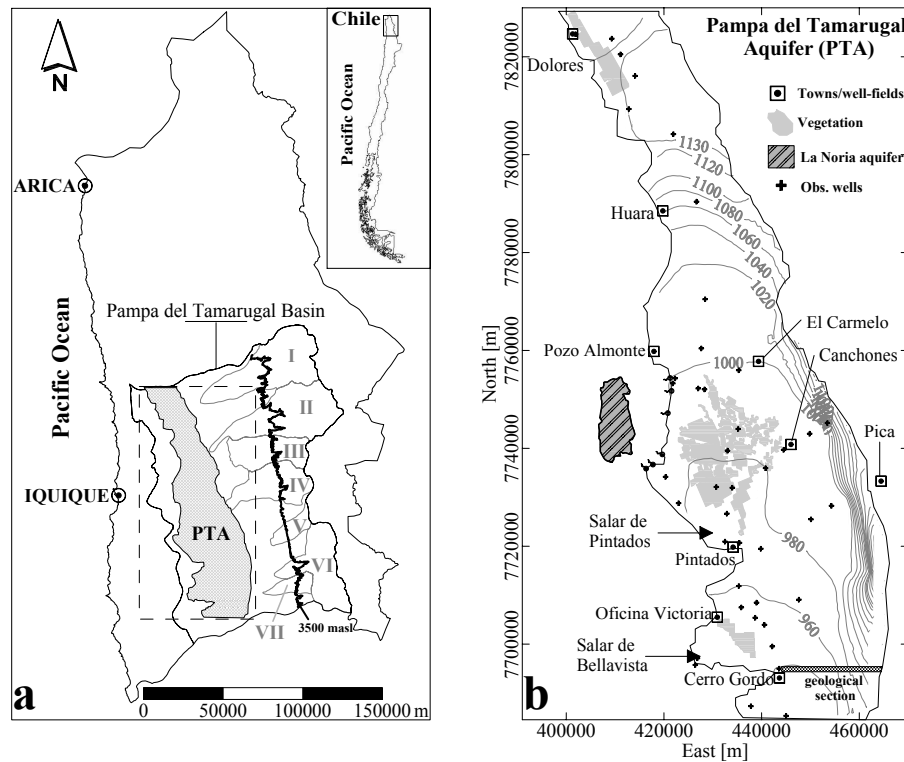


Fig. 1. (a) Location of study area, Pampa del Tamarugal Aquifer (PTA), eastern sub-basins (I: Aroma, II: Tarapaca, III: Quipisca, IV: Sagasca, V: Quisma, VI: Chacarilla, VII: Ramada) and topographic line 3500 m a.s.l. for the occurrence of precipitation (solid line), (b) groundwater elevation map (m a.s.l.) for the PTA for year 1960 (assumed steady-state conditions) (adapted from Rojas and Dassargues, 2007).

The study of the PTA has been a concern since the early 1960s and many local institutions, international aid agencies, and private companies have attempted to describe the functioning of this regional aquifer and its interactions with the scarce surface water (see e.g. Fritz et al., 1981; Suzuki and Aravena, 1985; DGA, 1988, 1996; Magaritz et al., 1990; DICTUC, 1995, 2005, 2007, 2008; JICA-DGA-PCI, 1995; Aravena, 1995; Salazar et al., 1999; DSM, 2002; Houston, 2002; Risacher et al., 2003; Rojas and Dassargues, 2007). These studies have been motivated by an increasing need to secure groundwater resources for drinking water supply, agricultural activities and, to a greater extent, by the growing pressure to provide fresh groundwater for the development of mining activities. As a result of these studies, several groundwater flow models (based on different conceptual models), ranging from simple approximations to more elaborated models, have been developed to simulate the behaviour of the PTA under different stress conditions.

DGA (1988) reported one of the first efforts to model the groundwater flow in the PTA. A one-layer model assuming steady-state conditions for the year 1960 and a transient-state model for the period 1960–1986 were developed, both using a zonation approach to describe the hydraulic conductivity field. These models satisfactorily reproduced the general flow pattern of the aquifer. Groundwater recharge coming

from the most northern sub-basin (I: Aroma) (Fig. 1a), however, was neglected because the simulated values did not satisfactorily reproduce the observed groundwater heads in the northern part of the modelled domain. As explained latter, the definition of the recharge mechanism will play an important role in the uncertainty assessment for the PTA. A second model (two-layer) was developed by JICA-DGA-PCI (1995) assuming steady-state conditions for year 1993 and using constant values for hydraulic conductivities for each layer. This model satisfactorily reproduced the flow patterns of the aquifer and the global water balance. Calibration of this model, however, assumed an additional recharge mechanism where a significant amount of groundwater recharge originated from a system of faults and deep fissures connected with Altiplano aquifers. This recharge mechanism was based on results that demonstrated the presence of fresh and recent groundwater at shallow levels in the centre part of the PTA (Magaritz et al., 1990). Recent studies, however, suggest that recharge is taking place as a result of infiltrating runoff in the apex of the alluvial fans originating from the eastern sub-basins due to flash flood events (Houston, 2002). The latter could also potentially explain the presence of fresh and recent groundwater in the centre part of the PTA, hence, contradicting the recharge mechanism assumed by JICA-DGA-PCI (1995). By neglecting this recharge mechanism a substantial

difference in the alternative conceptualizations used to model the PTA was established. In addition, steady-state conditions used in the model of JICA-DGA-PCI (1995) are rather questionable as suggested by present-day data (see e.g. Rojas and Dassargues, 2007). A third model (two-layer) was developed by DSM (2002). Similarly to the model of JICA-DGA-PCI (1995), this model included the groundwater recharge originating from deep fissures connected with Altiplano aquifers, however, steady-state conditions were assumed for year 1960 and a zonation approach for each layer was used to describe the hydraulic conductivity field. This model satisfactorily reproduced the overall flow field and water balance of the aquifer. Rojas and Dassargues (2007) developed a fourth model (one-layer) aimed at updating the knowledge of the system, extending the transient calibration to more recent data (1960–2004), and acting as a middle-point between models previously created. This model assumed steady-state conditions for year 1960 and *did not include* the additional groundwater recharge from deep fissures in the development of the conceptual model. As models previously developed, the model of Rojas and Dassargues (2007) acceptably reproduced the global water balance and the flow patterns of the PTA. Finally, two additional groundwater flow models for the PTA have been developed by DICTUC (2005, 2008). These models were successfully calibrated and correctly reproduced the water balance and the general flow patterns of the PTA, although the assumed steady-state conditions (for years 1960 and 1987) and the approach to model the spatial distribution of the hydraulic conductivity (zonation versus a Random Space Function-RSF approach) were rather different.

It is important to note, with the exception of the work of Rojas and Dassargues (2007), that groundwater flow models previously described lack a consistent uncertainty analysis and the corresponding authors assumed an accurate and unique description of the input (forcing) data, parameters, and conceptual model, hence, neglecting uncertainty in these terms.

To what extent any of these models can satisfactorily be used to manage the groundwater resources of the PTA and to predict its responses to future stress conditions is clearly debatable. These models were successfully calibrated using a unique and limited set of observed groundwater heads (between 40 and 55 observations) and rough estimations for the global water balance (e.g. evaporation from salar-playas and transpiration from forested areas) of the aquifer; hence, they are all valid representations of the groundwater system. They consider, however, different recharge mechanisms, geological interpretations, modelling of the observed heterogeneities in the hydraulic conductivity field, definition of the boundary conditions, surface extensions, temporal scales (time-steps), spatial scales (grid-sizes), steady-state initial conditions, and numerical approaches to discretise the model domain. Due to the differences in model conceptualization individual model predictions are prone to bias and

possibly inconsistent and conflicting when the results of the alternative conceptual models are compared. Additionally, predictive uncertainty estimations based on a single member of the ensemble of models previously described are likely to be under-dispersive due to omitting (potentially feasible) alternative conceptualizations from the analysis. These points are critical for the sustainable management of the PTA where human pressure for newly developed freshwater resources is considerably high and the uncertainty due to climate conditions is relatively important.

It has been recently suggested that uncertainties in groundwater model predictions are largely dominated by this type of uncertainty (conceptual model uncertainty) and that accounting for parametric uncertainty solely does not allow compensating for conceptual model uncertainty in model predictions (Bredehoeft, 2003; Neuman and Wierenga, 2003; Neuman, 2003; Ye et al., 2004; Højberg and Refsgaard, 2005; Poeter and Anderson, 2005; Bredehoeft, 2005; Refsgaard et al., 2006, 2007; Meyer et al., 2007; Seifert et al., 2008; Rojas et al., 2008). This is especially important for the case when predicted variables are not included in the data used for calibration (Højberg and Refsgaard, 2005; Troldborg et al., 2007). The latter is certainly the case for PTA where only a scarce set of groundwater heads has been regularly used to calibrate the groundwater flow models whereas predictions are made for flow components and groundwater fluxes. The impossibility to include flow-related observations due to the arid nature of the study area (which is located in the Atacama's desert) has made model calibration even more challenging. Under these conditions, non-uniqueness of parameters and equifinality problems may have a significant impact on models' results.

Rather than relying on a single conceptual model of a given groundwater system, it seems more appropriate to consider a range of plausible system representations and analyse the combined multi-model output to assess the predictive uncertainty (e.g. Harrar et al., 2003; Højberg and Refsgaard, 2005; Poeter and Anderson, 2005; Meyer et al., 2007; Rojas et al., 2008; Ijiri et al., 2009). Whereas predictions based on a single conceptualization are more likely to be biased and under-dispersive, estimates based on an ensemble of models are less (artificially) conservative and are more likely to capture the unknown true predicted value (Neuman, 2003; Rojas et al., 2008).

One approach to deal with this type of uncertainty has been recently proposed by Rojas et al. (2008, 2009b). The proposed methodology accounts for predictive uncertainty arising from inputs (forcing data), parameters, alternative conceptual models, and (potentially) the definition of alternative scenarios by combining Generalized Likelihood Uncertainty Estimation (GLUE) (Beven and Binley, 1992) and Bayesian Model Averaging (BMA) (Draper, 1995; Kass and Raftery, 1995; Hoeting et al., 1999). The key idea behind this approach is the concept of equifinality, that is, many combinations of model structures and parameter sets may

Table 1. Summary of alternative conceptual models.

Geological setup and spatial distribution of K	Recharge mechanism	
	Eastern sub-basins	Eastern sub-basins+ deep fissures in basement rocks
Two-layer and constant value of K for each layer	M1a	M1b
One-layer and zonation of K (22 zones)	M2a	M2b
One-layer and (unconditional) RSF representing K	M3a	M3b
One-layer and (conditional) RSF representing K	M4a	M4b

provide (equally) good reproductions of the observed system response (Beven, 2006). Using a hypothetical setup, Rojas et al. (2008) explored the global likelihood response surface of all possible combinations of plausible model structures, forcing data and parameter values in order to select those simulators that perform well. For each model structure, the posterior model probability (model weights to aggregate multi-model predictions) was obtained by integrating the likelihood measures over the retained simulators for that model structure. The posterior model probabilities were subsequently used in BMA to weight the predictions of the competing models when assessing the joint predictive uncertainty. Important aspects of this method are that (1) it does not rely on a unique optimal parameter set for each conceptual model to assess the joint predictive uncertainty, thus, avoiding compensation of conceptual model errors due to biased parameter estimates; (2) weights for combining model predictions are obtained considering the full sampled space; (3) there is no implicit assumption about the conditional pdf's obtained for each alternative conceptualization; and (4) allows for the inclusion of prior information about conceptual models, inputs and parameters, and state variables to perform conditional simulations. The latter may potentially be used for penalizing models with too many parameters with respect to the number of observations, i.e. to comply with the principle of parsimony. A more detailed description of such priors or the methods to define them, however, are beyond the scope of this article. A complete description of the methodology and potential advantages are discussed in Rojas et al. (2008a; 2009b).

In this work we assess the uncertainty in groundwater flow modelling of PTA in a fully integrated GLUE-BMA approach accounting for uncertainties in input (forcing) data, parameters, and conceptual models. Two additional aims of this work are to demonstrate the applicability of the method of Rojas et al. (2008) to regional-scale aquifer systems and to illustrate the effects of neglecting conceptual model uncertainty in a real-world modelling exercise. We propose an ensemble M of eight alternative conceptualizations covering all major features of groundwater flow models (independently) developed until present for the PTA. Proposed conceptualizations range from simple one- and two-layer approximations

to more complex models where the spatial distribution of the hydraulic conductivity field follows the theory of Random Space Functions (RSF). We consider two mechanisms to describe the recharge process: one considering only recharge due to groundwater flows originating from the eastern sub-basins (Fig. 1b), and the other (complementary to the previous one) due to a system of faults and deep fissures connecting the PTA with Altiplano aquifers (see Table 1). Based on the results of Rojas et al. (2009a), we consider spatial conditioning of the hydraulic conductivity field to a set of available hydraulic conductivity measurements. Finally, we analyse the full predictive uncertainty on groundwater heads, groundwater balance components and groundwater fluxes following the approach described in Rojas et al. (2008).

We must emphasize that it is not the aim of this work to select “the best model” out of an ensemble of model candidates, i.e. solve a model selection problem, rather the objective is to solve a predictive multi-model problem.

The remainder of this paper is organized as follows. In Sect. 2, we provide a condensed overview of GLUE and BMA followed by a description of the procedure to integrate these methods. Section 3 details the study area where the integrated uncertainty assessment methodology is applied. Implementation details such as the different conceptualizations, sampling of parameters and the summary of the modelling procedure are described in Sect. 4. Results are discussed in Sect. 5 and a summary of conclusions is presented in Sect. 6.

2 Integrated uncertainty assessment methodology

Sections 2.1 and 2.2 summarize the basis of GLUE and BMA methodologies, respectively, for more details the reader is referred to Rojas et al. (2008, 2009b). Section 2.3 presents the description on how to integrate these methodologies.

2.1 Generalized likelihood uncertainty estimation (GLUE) methodology

Being a Monte Carlo simulation technique relying on the concept of equifinality (Beven and Freer, 2001), GLUE rejects the idea of a single correct representation of a system in favour of many acceptable system representations (Beven,

2006). For each potential system simulator, sampled from a prior set of possible system representations, a likelihood measure (e.g. Gaussian, trapezoidal, model efficiency, inverse error variance) is calculated, which reflects its ability to simulate the system responses, given the available observed dataset \mathbf{D} . Simulators that perform below a subjectively defined rejection criterion are discarded from further analysis and likelihood measures of retained simulators are rescaled so as to render the cumulative likelihood equal to one. Ensemble predictions are based on the predictions of the retained set of simulators, weighted by their respective rescaled likelihood.

Following the notation of Rojas et al. (2008), let us consider a set of plausible model structures contained in the ensemble $\mathbf{M} = \{M_1, M_2, \dots, M_k, \dots, M_K | K < \infty\}$, a set of parameter vectors $\Theta = (\theta_1, \theta_2, \dots, \theta_l, \dots, \theta_L)$ and a set of input variable vectors $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_m, \dots, \mathbf{Y}_M)$, and denote the observed and simulated system variable vectors as $\mathbf{D} = (D_1, D_2, \dots, D_n, \dots, D_N)$ and $\mathbf{D}^* = (D_1^*, D_2^*, \dots, D_n^*, \dots, D_N^*)$, respectively.

Then, $L(M_k, \theta_l, \mathbf{Y}_m | \mathbf{D})$ represents the likelihood of the k -th model structure (M_k) parametrized with l -th parameter vector (θ_l) and forced by m -th input data vector (\mathbf{Y}_m) to represent the true system, given the observations in \mathbf{D} . Rojas et al. (2008) observed no significant differences in the estimation of posterior model probabilities, predictive capacity, and conceptual model uncertainty when a Gaussian, a model efficiency based, or a Fuzzy-type likelihood function was used. The analysis in this work is therefore confined to a Gaussian likelihood function.

2.2 Bayesian model averaging (BMA)

BMA is a statistical procedure that provides a coherent framework for combining predictions from multiple competing conceptual models to attain a more realistic and reliable description of the predictive uncertainty. BMA infers average predictions by weighing individual (conceptual) model predictions based on their relative skill, with predictions from better performing models receiving higher weights than those of worse performing models. Thus, BMA avoids having to choose a model over the others, instead, competing models are assigned different weights based on the observed dataset \mathbf{D} (Wasserman, 2000).

Following the notation of Hoeting et al. (1999), if Δ is a quantity to be predicted, the full BMA predictive distribution of Δ for a set of alternative conceptual models defined by $\mathbf{M} = (M_1, M_2, \dots, M_k)$, is given by (Draper, 1995; Rojas et al., 2008).

$$p(\Delta | \mathbf{D}) = \sum_{k=1}^K p(\Delta | \mathbf{D}, M_k) p(M_k | \mathbf{D}). \quad (1)$$

Equation (1) is an average of the predictive distributions of Δ under each alternative conceptual model, $p(\Delta | \mathbf{D}, M_k)$,

weighted by their posterior model probability, $p(M_k | \mathbf{D})$. The posterior model probabilities reflect how well model M_k fits the observed data \mathbf{D} and can be computed using Bayes' rule (see e.g. Rojas et al., 2008)

$$p(M_k | \mathbf{D}) = \frac{p(\mathbf{D} | M_k) p(M_k)}{\sum_{l=1}^K p(\mathbf{D} | M_l) p(M_l)} \quad (2)$$

where $p(M_k)$ is the prior probability of model M_k , and $p(\mathbf{D} | M_k)$ is the integrated likelihood of model M_k .

The leading moments of the full BMA prediction of Δ are given by (Draper, 1995; Rojas et al., 2008)

$$E[\Delta | \mathbf{D}] = E_M[E(\Delta | \mathbf{D}, M)] = \sum_{k=1}^K E[\Delta | \mathbf{D}, M_k] p(M_k | \mathbf{D}) \quad (3)$$

$$\begin{aligned} \text{Var}[\Delta | \mathbf{D}] &= E_M[\text{Var}(\Delta | \mathbf{D}, M)] + \text{Var}_M[E(\Delta | \mathbf{D}, M)] \\ &= \sum_{k=1}^K \text{Var}[\Delta | \mathbf{D}, M_k] p(M_k | \mathbf{D}) + \sum_{k=1}^K (E[\Delta | \mathbf{D}, M_k] - E[\Delta | \mathbf{D}])^2 p(M_k | \mathbf{D}). \end{aligned} \quad (4)$$

From Eq. (4) the variance of Δ consists of two terms: the first representing the within-models variance, and the second representing the between-models variance (Rojas et al., 2008).

2.3 Multi-model approach to account for conceptual model and scenario uncertainties

Combining GLUE and BMA to account for conceptual model uncertainties involves the following sequence of steps

1. A suite of alternative conceptualizations is proposed and prior model probabilities are assigned. This can be done on the basis of prior and expert knowledge about the site (see e.g. Meyer et al., 2007; Ye et al., 2008b; Rojas et al., 2009b). Should this prior knowledge reflects the soundness and plausibility of the alternative conceptualizations, prior model probabilities could potentially be used to penalize models with a high number of parameters, i.e. to fully comply with the principle of parsimony.
2. Prior ranges are defined for the input and parameter vectors under each plausible model structure. To keep the analysis at a neutral level, multi-uniform prior distributions could be assumed to perform the sampling of input and parameter vectors. Should sound and proper prior knowledge were available about input and parameter vectors, an attempt to include it in the form of non-uniform prior distributions should be made as far as possible (Ghosh et al., 2006).
3. A likelihood measure to assess model performance and a rejection criterion are defined. The latter can be based on exploratory runs (e.g. Rojas et al., 2008, 2009c), subjectively chosen threshold limits (e.g. Feyen et al., 2001) or set as a minimum level of performance (e.g. Binley and Beven, 2003).

4. For the suite of alternative conceptual models, input and parameter values are sampled using the Metropolis-Hastings (M-H) algorithm (Metropolis et al., 1953; Hastings, 1970; Chib and Greenberg, 1995; Gilks et al., 1995) to generate simulators of the system.
5. A value for the likelihood measure $p(\mathbf{D}|M_k, \theta_l) \approx L(M_k, \theta_l, \mathbf{Y}_m|\mathbf{D})$ is calculated for each simulator and, based on the rejection criteria, it is added to the subset A_k of retained simulators for model M_k or it is discarded by setting its likelihood to zero.
6. Steps 4–5 are repeated until the hyperspace of possible simulators is adequately sampled, i.e. when for each model M_k the first two moments of the conditional distributions of predicted state variables based on the retained likelihood weighted simulators converge to stable values, and the *R*-score (Gelman et al., 2004) for parameters and variables of interest converges to values close to one. The *R*-score expresses the ratio of within- to between-chains variability and, thus, approximate convergence of the M-H algorithm is diagnosed when the variability between chains is not larger than that within chains (Sorensen and Gianola, 2002).
7. The integrated likelihood of each model M_k is approximated by summing the likelihood weights of the retained simulators in the subset A_k , that is, $p(\mathbf{D}|M_k) \approx \sum_{l, m \in A_k} L(M_k, \theta_l, \mathbf{Y}_m|\mathbf{D})$ (Rojas et al., 2008).
8. By normalizing the integrated model likelihoods over the whole ensemble \mathbf{M} such that they sum up to one (using Eq. 2), the posterior model probabilities are obtained.
9. After normalization of the likelihood weighted predictions under each individual model (such that the cumulative likelihood under each model equals one), a multi-model prediction is obtained with Eq. (1). The leading moments of this distribution (considering the whole ensemble \mathbf{M}) are then obtained using Eqs. (3) and (4).

In addition, posterior model probabilities obtained in step (8) could potentially be used in the prediction stage of the alternative conceptual models under alternative scenarios (e.g. Rojas et al., 2009c).

Posterior model weights implicitly account for the number of parameters through the likelihood function. It is likely that models with more parameters will have higher likelihoods, as the latter are obtained from model fit solely. If both prior model probabilities and integrated model likelihoods (Eq. 2) are equal for alternative conceptual models, these conceptualizations will receive the same posterior model weight (independently of the number of parameters). The latter means that the evidence provided by the data did not support any

single model compared to the others. We acknowledge, however, that the GLUE-BMA method does not penalize more complex models, i.e. models with a higher number of parameters to comply with the principle of parsimony, through the model likelihood. Nevertheless, penalizing more complex models can be achieved by defining non-uniform prior model probabilities. These non-uniform prior distributions could penalize alternative conceptualizations on the basis of the number of parameters, complexity, plausibility or any other criteria adopted by the analyst. An analysis of this type, however, is beyond the scope of this article.

A second alternative to penalize models with a high number of parameters is to use model selection criteria (e.g. AIC, AICc, BIC or KIC) to approximate the posterior model probabilities. Although Ye et al. (2008a) have presented an insightful discussion about merits and demerits of alternative model selection criteria in multi-model applications, the debate on using one criterion over the others is far from being settled yet. In addition, in a recent work (Rojas et al., 2009c) we have shown that working with different model selection criteria to approximate the posterior model probabilities might lead to conflicting and misleading results in multi-model applications.

3 Study area

3.1 General description

The PTA is limited in the west by the coastal range and in the east by the Chilean pre-cordillera. It covers an area of ca. 5000 km² with dimensions of almost 160 km long, width between 20–60 km, and an average elevation of 1000 m a.s.l. (Fig. 1).

Recharge through direct precipitation on the PTA is not taking place as rainfall is negligible in the area. The eastern sub-basins (Fig. 1a), on the other hand, receive recharge from precipitation coming from high altitudes in the east. These sub-basins lie in a well-developed rain-shadow, and as a result there is a rapid decrease in rainfall as air masses move west and descend (Aravena et al., 1989, 1999; Houston, 2002, 2006). For an average hydrologic year surface watercourses disappear before reaching most of the alluvial fans located in the PTA suggesting that the aquifer is being recharged by part of this water through infiltration and lateral groundwater flows (Aravena, 1995). Extreme rainfall events, on the other hand, play an important role on the recharge mechanisms making the (overused) assumption of “average” recharge conditions rather questionable (Houston, 2006). In this regard, evidence of recent groundwater recharge due to flash flood events has been reported for the Chacarilla sub-basin (VI in Fig. 1a) by Houston (2002).

The PTA is located in the Atacama Desert and thus arid conditions are extreme with potential evaporation rates ranging between 2000 mm yr⁻¹ and 2500 mm yr⁻¹ (Rojas and

Dassargues, 2007). As a consequence, the presence of natural vegetation is limited to few places in the study area. These places correspond to natural or reforested areas located in Dolores, Salar de Pintados and Salar de Bellavista (Fig. 1b) (Rojas and Dassargues, 2007). These zones are composed of trees highly adapted to arid and saline conditions and, thus, they can sustain themselves by directly extracting groundwater with massive and well-developed root systems (FAO, 1989). Therefore, vegetative transpiration from these zones is an important component of the global water balance for the PTA.

3.2 Geology

The geology of the study area has been described in previous studies (e.g. Dingman and Galli, 1965; DGA, 1988; JICA-DGA-PCI, 1995; Digert et al., 2003; DICTUC, 2007). As described in Houston (2002), the basin is a complex asymmetric graben bounded in the west and in the east by N–S regional fault zones which started to form in the early Oligocene. Coarse conglomerates and gravels of the Sical and Altos de Pica Formations, eroded from the adjacent uplifted Coastal Range and Pre-Cordillera, comprise the lowermost sediments. Coarse clastic sediments continued to be deposited over wide areas throughout the Miocene. Some events of volcanic activity, from the eruptive centres located on the east, took place producing andesitic tuffs and ignimbrites. Towards the end of the Miocene a series of large alluvial fans began to develop. Since the Pliocene only minor alluvial and evaporitic sediments have been deposited in the basin.

Figure 2 shows a longitudinal geological profile (XX') of the study area and a plan view of the aquifer boundaries. In Fig. 2a upliftings of the basement rocks (Longacho Formation) are observed in the north and south limits. Also in the west (Coastal Range) and in the east, outcroppings of this formation are observed (Fig. 2b). Overlying the Longacho Formation is the Altos de Pica Formation, which is differentiated into lower and upper layers. In the uppermost strata, recent sediments mainly composed of saline alluvial deposits, gravel, sand and clay have been deposited (JICA-DGA-PCI, 1995).

3.3 Hydrogeology

The main aquifer system of the PTA is contained in units Q4 and Q3 (Fig. 2a) (DGA, 1988; JICA-DGA-PCI, 1995). Unit Q3 is composed of sand and gravel and is underlain by a thick clayey layer (Q2). Unit Q4 consists of sand and gravel with mud, and/or intercalated with mud layers and is overlain by unit Q3 from Huara to Salar de Bellavista area (Rojas and Dassargues, 2007). A relatively good qualitative data base is available for the description of these units. Hard data (hydraulic conductivity measurements, transmissivity, storativity), however, are scarce given the dimensions of the PTA

Table 2. Summary of aquifer parameters of the Pampa del Tamarugal Aquifer (PTA) obtained from 62 measurement points. Adapted from JICA-DGA-PCI (1995) and DICTUC (2008).

Zone	Transmissivity ($\text{m}^2 \text{d}^{-1}$)	Hydraulic conductivity (m d^{-1})	Storativity (–)
Dolores	7–905	1.1–35.9	3.0×10^{-4} –0.30
Huara	8–935	0.1–29.1	5.7×10^{-7} –0.08
Pozo Almonte – Pintados	21–2097	0.3–42.4	2.0×10^{-6} –0.20
Oficina Victoria – Cerro Gordo	29–1500	0.8–150	3.0×10^{-7} –0.33

and they are in the limit to perform a meaningful geostatistical analysis. A summary of the aquifer parameters for the PTA is presented in Table 2.

Figure 1b shows the groundwater elevation map for the year 1960 which is obtained from nearly 60 measurement points. The main groundwater flow direction is from north to south with an east-west component in the area of Pica. A groundwater divide is observed in the north area with part of the groundwater flowing towards the forested area in Dolores (Rojas and Dassargues, 2007). Hydraulic gradients are considerably steeper near Huara which is explained by the lower hydraulic conductivity of the deposits of Aroma and Tarapaca sub-basins (see Fig. 1a) (DGA, 1988). In the centre area of PTA, groundwater flows from east to west towards the forested Tamarugo areas and the Salar de Pintados. Both correspond to the main natural discharge areas of the aquifer for the year 1960. In the southern area (Oficina Victoria-Cerro Gordo) groundwater flow is mainly directed to the south-west (Rojas and Dassargues, 2007).

There is a general consensus that recharge to the PTA occurs mainly by groundwater flows from the eastern sub-basins, with estimations ranging between $76\,032 \text{ m}^3 \text{ d}^{-1}$ and $89\,510 \text{ m}^3 \text{ d}^{-1}$ (DGA, 1988; JICA-DGA-PCI, 1995; Aravena, 1995; Houston, 2002; Rojas and Dassargues, 2007; DICTUC, 2008). As discussed earlier, a mechanism where a substantial amount of groundwater recharge originating from a system of faults and deep fissures connected with Altiplano aquifers is also assumed by some authors (e.g. Magaritz et al., 1990; JICA-DGA-PCI, 1995; DSM, 2002). The magnitude of this recharge ranges between $11\,834 \text{ m}^3 \text{ d}^{-1}$ (DGA, 1996) and $24\,970 \text{ m}^3 \text{ d}^{-1}$ (JICA-DGA-PCI, 1995) and, until today, the validity and quantification of this recharge mechanism is an important source of uncertainty to develop a coherent conceptual model for the PTA.

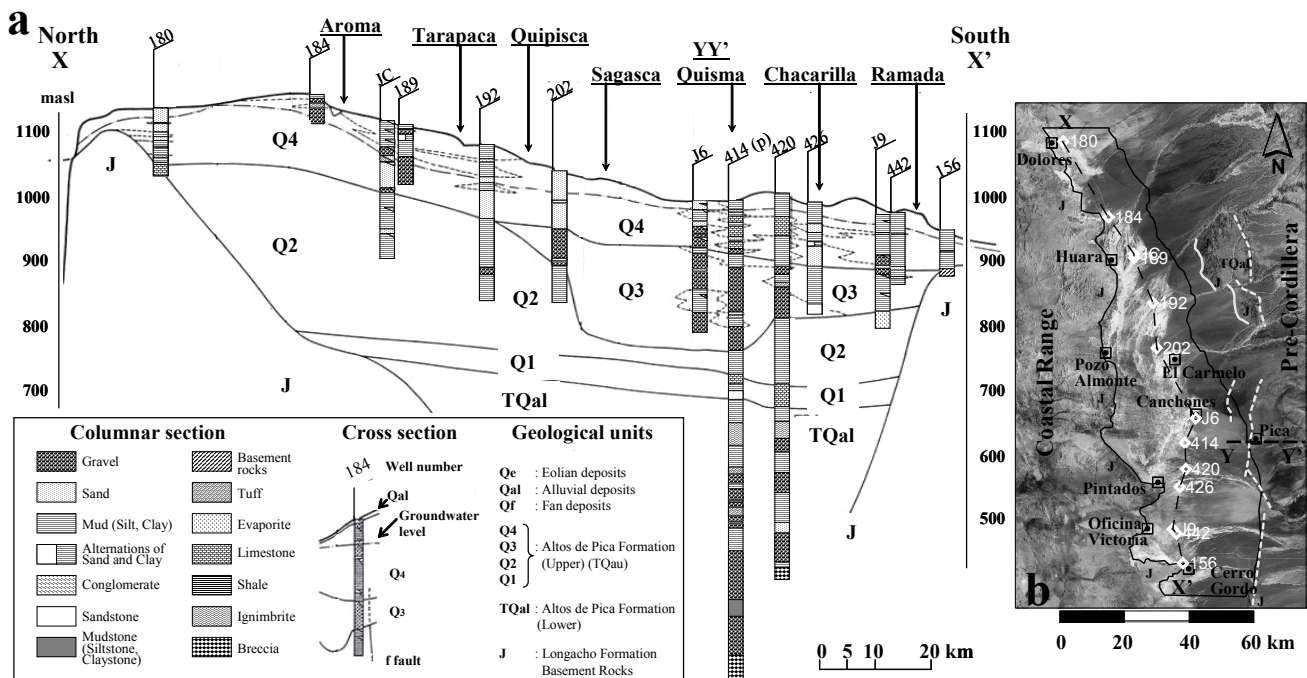


Fig. 2. (a) Geological longitudinal profile (XX') of PTA and (b) planar delimitation of the aquifer boundaries and basement rocks outcroppings (after Rojas and Dassargues, 2007).

Estimations of transpiration rates from forested areas (Fig. 1b) for the year 1960 range between ca. $1728 \text{ m}^3 \text{ d}^{-1}$ (DICTUC, 2008) and $18\,144 \text{ m}^3 \text{ d}^{-1}$ (DGA, 1988). Direct evaporation of groundwater from salares constitutes other important component of the water balance. In the study area, an active salar system exists comprising the Salar de Pintados and Salar de Bellavista (Fig. 1b) (see e.g. Risacher et al., 2003). An estimated evaporation of $46\,829 \text{ m}^3 \text{ d}^{-1}$ is obtained from DGA (1988) for the year 1960, whereas Rojas and Dassargues (2007) estimated a range between $35\,424 \text{ m}^3 \text{ d}^{-1}$ and $52\,013 \text{ m}^3 \text{ d}^{-1}$ following the calibration of a steady-state model for the same year.

Groundwater outflows from the PTA are estimated in the order of ca. $1728 \text{ m}^3 \text{ d}^{-1}$ in the north boundary, and ca. $8640 \text{ m}^3 \text{ d}^{-1}$ in the geological section defined between the outcrops located in Cerro Gordo and the east limit of the model (Fig. 1b) (DICTUC, 2008).

Recently, a hydrogeologic connection with a local aquifer termed La Noria has been suggested by DICTUC (2008) (Fig. 1b). It is proposed that the PTA is recharging this local aquifer through a local system of faults and the estimated groundwater recharge flows for the present situation vary between $1555 \text{ m}^3 \text{ d}^{-1}$ and $4320 \text{ m}^3 \text{ d}^{-1}$.

4 Implementing the multi-model approach

4.1 Alternative conceptual models

An ensemble M including eight alternative conceptualizations was considered to describe the PTA. These conceptual models aimed at covering the main features of all models previously developed (Fig. 3). All members of M comprised six common elements: boundary conditions expressed as constant heads at the north and south limits, hydrogeological connection with La Noria aquifer, a transpiration zone located in Pintados, an evaporation zone accounting for Salar de Pintados and Salar de Bellavista, and assumed steady-state conditions for year 1960. The distinctive elements among models were the number of layers, the representation of the hydraulic conductivity field and the recharge mechanisms (see Table 1). Model M1 (Fig. 3a) considered a two-layer system explicitly accounting for units Q3 and Q4 (see Fig. 2a), whereas models M2 (Fig. 3b), M3 (Fig. 3c) and M4 (Fig. 3d) considered a one-layer system. Model M1 considered constant values of hydraulic conductivity for each layer, model M2 included a spatial zonation approach obtained from Rojas and Dassargues (2007), and models M3 and M4 used the theory of Random Space Functions (RSF) to describe the hydraulic conductivity field, with model M4 conditioning the hydraulic conductivity realizations on available hydraulic conductivity measurements. To obtain the

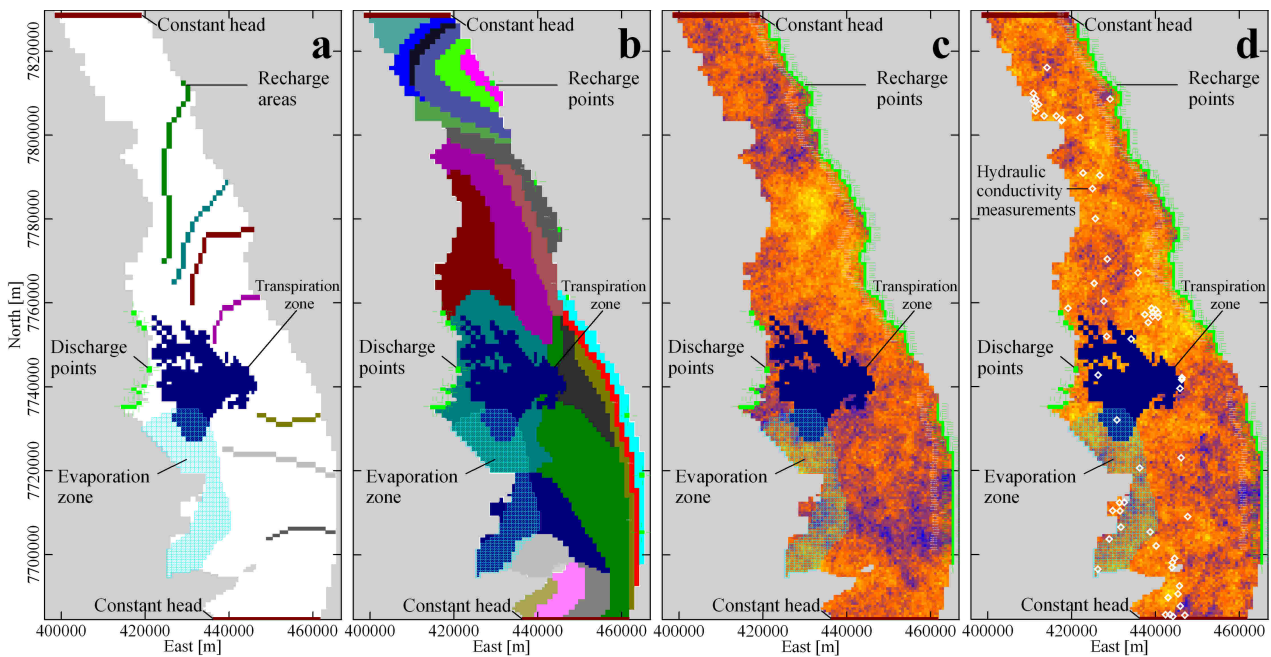


Fig. 3. Model setup for four alternative conceptual models: (a) M1 (two-layer model with constant values of K in each layer), (b) M2 (one-layer model with zonation of K), (c) M3 (one-layer model, unconditional RSF), and (d) M4 (one-layer model, conditional RSF).

conditional realizations of the K -fields, we employed “conditioning by kriging” (Chilès and Delfiner, 1999). Figure 4 shows the histogram for $\ln K$ and the modelled variogram used to simulate the hydraulic conductivity fields in the frame of models M3 and M4. The fitted model corresponded to a nested structure including a nugget effect of 0.6 and an exponential variogram with a practical range of 10.2 km and a sill (contribution) of 1.6. For the recharge inflows originating from the eastern sub-basins, model M1 considered areal recharge rates distributed over small areas of the alluvial fans (Fig. 3a), model M2 included point recharge fluxes in the apex of the alluvial fans (Fig. 3b), and models M3 and M4 considered recharge fluxes distributed over long sections of the eastern boundary (Fig. 3c and d).

In addition, each of the four models depicted in Fig. 3 considered an alternative “b” version where the recharge mechanism that assumes a connection with Altiplano aquifers was included, hence, adding to the groundwater recharge flows from the eastern sub-basins. These recharge rates were spatially distributed over the entire active model domain. In summary, eight alternative conceptualizations were defined to analyse the combined uncertainty arising from inputs, parameters, and conceptual models.

4.2 Prior distributions

The common elements to the alternative conceptualizations defined a common group of eight parameters, namely, the elevations of the constant head cells at the north (CH.N)

and south (CH.S) limits, recharge inflows from the eastern sub-basins (RECH), recharge inflows from faults and deep fissures (RECH.BAS) (only for “b” version models), transpiration outflows (TRANSP), evaporation rates from salares (EVTR), extinction depth of the evaporation process (EXTD), and outflows to La Noria aquifer (NORIA). In addition, for models M1 and M2, two and twenty-two additional parameters representing the hydraulic conductivity values for each layer or zones were included, respectively. In a previous work, Rojas and Dassargues (2007) did an extensive sensitivity analysis for a model analogous to model M2 used in this article. From that analysis, the most sensitive parameters were recharge rates and hydraulic conductivities, which showed some degree of correlation given the steady-state nature of the model. Elevation of the south boundary condition showed moderate sensitivity whereas parameters related to the evaporation process were relatively insensitive. We acknowledge that parameter correlations may yield biased results, however, an analysis of the effects of these correlations on the GLUE-BMA results is beyond the scope of this article.

We assigned equal prior model probabilities (1/8) to the eight alternative conceptualizations and adopt uniform prior distributions for the unknown inputs and parameters. Using these priors, we expect that the information contained in the data, expressed by the likelihood function, should dominate the form of the resulting posterior distributions. The ranges that describe the prior uniform distributions of the unknown variables are presented in Table 3. Despite some parameter

Table 3. Range of prior uniform distributions for unknown parameters.

Parameter			Range	
			Minimum	Maximum
Recharge	(RECH)	($\text{m}^3 \text{d}^{-1}$)	0	345600
Recharge from deep fissures in basement rocks	(RECH_BAS)	($\text{m}^3 \text{d}^{-1}$)	0	172800
Transpiration forested areas	(TRANSP)	($\text{m}^3 \text{d}^{-1}$)	0	172800
Discharge to La Noria aquifer	(NORIA)	($\text{m}^3 \text{d}^{-1}$)	0	86400
Evaporation rate	(EVAP)	(md^{-1})	0	0.01
Extinction depth	(EXTD)	(m)	0	20
Elevation constant head north	(CH_N)	(m)	1075	1120
Elevation constant head south	(CH_S)	(m)	875	920
Hydraulic conductivity ^a	(K)	(md^{-1})	0	100

^a Included only for models M1 and M2.

ranges shown in Table 3 span several orders of magnitude, samples were efficiently selected by the Metropolis-Hastings algorithm and clear zones of attraction were defined.

4.3 Simulation procedure

Simulation of steady-state flow for the year 1960 employed MODFLOW-2000 (Harbaugh et al., 2000). The modelled domain covered an area of ca. 4300 km² with a length of 160 km and between 18 km and 40 km wide. Uniform cell size of 600 m-600 m was used to discretise the modelled domain resulting in 242 rows and 118 columns. This cell size was defined on the basis of values used in previous studies (DGA, 1988; JICA-DGA-PCI, 1995; DSM, 2002; DICTUC, 2005; Rojas and Dassargues, 2007; DICTUC, 2008) and on pragmatic reasons to make the problem computationally tractable. It is worth mentioning, however, that preliminary runs were performed to estimate the computational time for different cell sizes. Models M1a and M1b (see Table 1) were defined as three-dimensional with varying cell thicknesses represented by stratigraphic units Q3 and Q4 (see Rojas and Dassargues, 2007). Models M2, M3 and M4 (versions a and b) (see Table 1), on the other hand, were defined as two-dimensional with a layer thickness varying between ca. 70 m and 300 m. The evaporation and the recharge packages of MODFLOW-2000 were used to represent the evaporation from salares and the transpiration from the forested areas, respectively. As discussed in Rojas and Dassargues (2007), surface extension of forested areas as well as forest species are well documented in previous studies and, thus, the estimation of past and present transpiration rates are relatively accurate. As a consequence, we opted for a head-independent boundary condition to represent the transpiration using the recharge package with negative recharge rates. Fixed inflow or outflow fluxes were represented with the well package.

A Gaussian likelihood measure was implemented to assess the model performance, i.e. to assess the ability of a simulator (conceptual model + set of parameters) to reproduce the observed dataset \mathbf{D} , which consisted of 42 observed heads (Fig. 1b). These heads were obtained for both units (Q3 and Q4) since observation wells were screened at multiple sections. The range of this measurements was between 915 m a.s.l. and 1033 m a.s.l. For convenience, we employed updated information about the observation wells obtained from DICTUC (2008) which is reported as the most accurate information until present. From exploratory runs and considering the dimensions of the modelled domain, a departure of $\pm 3\sigma_h$ m from the observed head in each observation well is defined as rejection criterion, where the standard deviation of observed heads (σ_h) was assumed as 10 m. This value was the basis for the rejection criterion implemented in the GLUE-BMA method and it was obtained from preliminary runs. The objective of these preliminary runs was to achieve a trade-off between computational time and number of “behavioural” simulations in the subset A_k . This ensured an efficient implementation of the sampling algorithm (Metropolis-Hastings) used in the methodology. Given the dimensions of the study area and the range of the head measurements (118 m) we considered this value as appropriate. In summary, if $h_{\text{obs}} - 30\text{m} < h_{\text{sim}} < h_{\text{obs}} + 30\text{m}$ a Gaussian likelihood measure is calculated, otherwise the likelihood is zero.

Sampling of parameters from the prior ranges presented in Table 3 was performed using the Metropolis-Hastings (M-H) algorithm (Metropolis et al., 1953; Hastings, 1970; Chib and Greenberg, 1995; Gilks et al., 1995). From exploratory runs, 100 parallel Markov chains starting from randomly selected points defined in the prior parameter ranges (Table 3) were implemented to proceed with the M-H algorithm for models M1 and M2. For models M3 and M4, sampled parameters were combined with hydraulic conductivity realizations generated using the spatial correlation structure defined in Fig. 4. Based on the results of Rojas et al. (2009a), an ensemble of

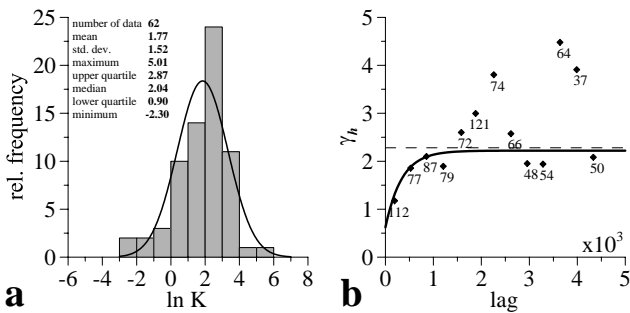


Fig. 4. (a) Histogram for hydraulic conductivity measurements and (b) modelled variogram to describe the spatial correlation structure of $\ln K$ for the PTA. Horizontal dashed line represents the variance of measurements.

100 hydraulic conductivity realizations was defined as minimum number to initiate the Markov chains. That is, when a random set of parameters combined with a hydraulic conductivity realization agreed with the rejection criterion, the hydraulic conductivity field was set as fixed and the Markov chain proceeded for the set of parameters solely. Alternatively, one could generate a large set of hydraulic conductivity realizations and develop a chain for each individual realization. However, given the high number of hydraulic conductivity realizations required to properly represent spatial variability, such approach is computationally still intractable.

Multivariate normal distributions centred on the previous parameter values were selected as proposal distribution, i.e. $q(\theta^*|\theta_{i-1}) \sim N(\theta_{i-1}|\Sigma_\theta)$, for models M1, M2, M3 and M4. Individual variance terms of Σ_θ were modified by trial-and-error until an acceptance rate in the order of 20–40% was achieved for successive steps of all Markov chains. For each proposed set of parameters a new Gaussian likelihood value was calculated in function of the agreement between observed and simulated groundwater heads at the 42 observation wells depicted in Fig. 1b. The mixing of the chains and the convergence of the posterior probability distributions was monitored using the R-score (Gelman et al., 2004) and the length of the *burn-in* samples was defined from visual inspection of the plotting series of exploratory runs for parameters and variables of concern.

To avoid negative effects of autocorrelation within successive steps of a chain (see e.g. Sorensen and Gianola, 2002), final chains were thinned before calculating summary statistics. Thinning an MCMC chain means that not all (chain) samples are recorded, instead samples are stored after k -th iterations. A thinned chain contains most of the information but it takes up less space in memory (Gilks et al., 1995). The resulting total parameter sample (after discarding the *burn-in* samples and after thinning the original sample) can be considered as a sample from the posterior distribution given the observed dataset \mathbf{D} for each alternative conceptual model. Using these discrete samples from the M-H algorithm, the

integrated likelihood of each conceptual model, $p(\mathbf{D}|M_k)$ in Eq. (2), is approximated by summing over all the retained likelihood values for model M_k . The posterior model probabilities are then obtained by normalizing over the whole ensemble M . For each series of predicted variables of concern, a cumulative predictive distribution, $p(\Delta|\mathbf{D}, M_k)$, is approximated by normalizing the retained likelihood values for each conceptual model such that they sum up to one. The leading moments of the full BMA predictive distribution accounting for input, parameter and conceptual model uncertainties are then obtained using Eqs. (3) and (4).

5 Results and discussion

5.1 Validation of the M-H algorithm results

Several aspects of the implementation of the M-H algorithm, such as the acceptance rate, the length of the *burn-in* samples and the proper mixing of the chains were checked to validate the results. Acceptance rates across all models ranged between 21% and 48%, values considered acceptable (see e.g. Gilks et al., 1995; Makowski et al., 2002). Based on the visual inspection of the plotting series for all parameters and variables of concern (e.g. recharge inflows, transpiration outflows, evaporation outflows, groundwater fluxes) and the values for the R-score of Gelman et al. (2004), the length of the *burn-in* samples was defined at $0.2N$ (with N being the total chain length). After discarding the *burn-in* samples, this resulted in chains of 20 000 elements. Across all models, the variation for the R-score for parameters of interest ranged between 1.003 and 1.107 for parameters CH_N and RECH_BAS, respectively. For variables of interest (groundwater flux at southern section) the variation for the R-score ranged between 1.001 and 1.110. For the problem at hand, and given the dimensions of the modeled domain and the complex interactions between parameters and variables of concern, we accepted proper mixing of the chains for values of the R-score close to 1.1, which is the value recommended for most problems by Gelman et al. (2004, p.297). As an example, Fig. 5 shows the results for model M2a for the key parameter recharge inflows. Figure 5a shows the development of fifty independent chains of 20 000 elements after discarding the initial *burn-in* samples. A good overlap of the chains is observed indicating proper mixing and convergence. Additionally, in order to get statistically independent results and given the autocorrelation induced by successive steps of the Markov chains, the original sample was thinned. We tried several thinning intervals (frequencies) before reaching a compromise between size of the chains and information content. For each of these trials we recomputed the first two moments to ensure consistency of the retained samples. Figure 5b shows the effect of thinning the original sample of 1000 000 elements after every 25 iterations. For the original sample the autocorrelation factor is highly

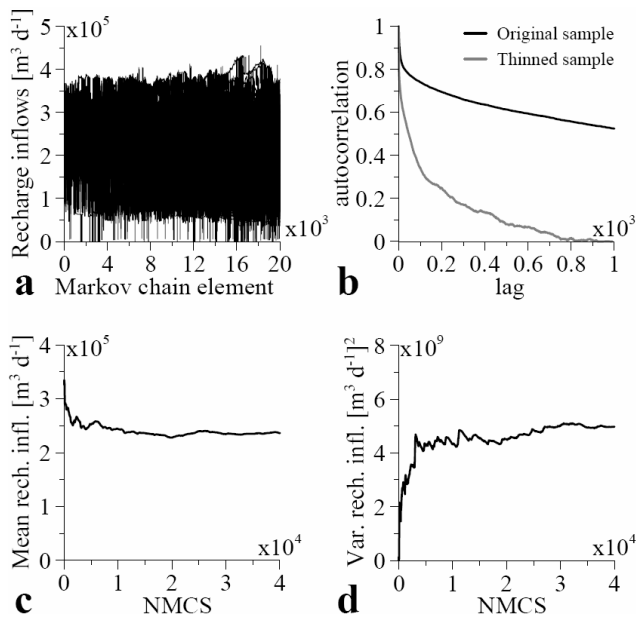


Fig. 5. Results from the M-H algorithm for recharge inflows from the eastern sub-basins for model M2a: (a) 100 independent Markov chains, (b) autocorrelation function for the original sample and thinned sample, (c) convergence of the mean for the predictive distribution of recharge inflows (thinned sample), and (d) convergence of the variance for the predictive distribution of recharge inflows (thinned sample).

persistent even for lags of 1000 whereas for the thinned sample autocorrelation is below ~ 0.3 for lags in the order of 150–200. Figure 5c and d shows a satisfactory convergence of the first two moments for the predictive distribution obtained with the thinned sample. Similar results were obtained for the other parameters and variables of interest. Therefore, the thinned parameter samples of 40 000 elements for each conceptual model were considered to be an independent sample from the target posterior distributions. These independent samples were combined in a final sample containing 320 000 (40 000 \cdot 8 models) elements which was used to calculate summary statistics of the posterior GLUE-BMA predictions.

5.2 Likelihood response surfaces

Figure 6 shows scatter plots of likelihood values for models M1a, M2a, M3a, and M4a for three key groundwater flow components, namely, recharge inflows, evaporation outflows, and groundwater fluxes passing through a section defined by the outcrop at Cerro Gordo and the east boundary of the modelled domain (see Fig. 1b). The latter has been estimated in previous studies, hence, is included here for comparison.

In general, the global likelihood response surfaces for models M3a and M4a are better defined and less disperse compared to models M1a and M2a, indicating the relevance

of describing the hydraulic conductivity field in more detail and the importance of conditioning the field to hydraulic conductivity measurements. This result is in full agreement with the findings of Rojas et al. (2009a). For recharge inflows (Fig. 6a–d), a clear region of attraction is identified, which contains the range of estimations made in previous studies. Conditioning the hydraulic conductivity field shows a clear effect on the peakedness of the likelihood surface, which is consistent with the range of previous estimations. The latter indicates the relevance of properly honouring the measured values of hydraulic conductivity. Evaporation outflows (Fig. 6e–h) shows a similar pattern as the one for recharge inflows, however, the peak of the likelihood surfaces defined by models M3a and M4a is somewhat biased compared to the range of previous estimations. An attraction zone for the likelihood response surface is observed, with the range of the previous estimations defining an upper limit for the evaporation outflows. Likewise recharge inflows, describing the hydraulic conductivity field in more detail decreases the spread of the likelihood surfaces, hence, decreasing uncertainty. For evaporation outflows under model M1a (Fig. 6e), and also groundwater outflows under model M2a (Fig. 6j), however, the range defined by previous estimations is located rather at the tail of the likelihood response surfaces, indicating that either these models fail to properly estimate these two flow components or previous estimations are rather conservative. For the groundwater outflows (Fig. 6i–l), increasing the complexity of the model considerably improves the accuracy of the estimation, suggesting that conditioning to hydraulic conductivity measurements plays an important role in the estimations of groundwater fluxes calculated at the Cerro Gordo section (Fig. 1b). Figure 6k and l shows that the likelihood response surfaces obtained with models M3a and M4a are consistent with previous estimations. When the hydraulic conductivity field is conditional on hydraulic conductivity measurements, however, previous estimations for the groundwater outflows at the southern section correspond to a local maximum in the likelihood response surface. From Fig. 6l, however, another highly likely value can be identified for these groundwater outflows in the range between 3310–4433 $\text{m}^3 \text{d}^{-1}$. This new range indicates an average reduction of the groundwater fluxes through the southern section of ca. 55% compared to previous estimations. This may play an important role in the management of the groundwater resources in the study area since groundwater fluxes through the section defined at Cerro Gordo are considered recharge inflows to the local aquifer system of Salar Viejo located south of PTA (DICTUC, 2008).

5.3 Posterior model probabilities

Table 4 shows the results for the integrated model likelihoods and the posterior model probabilities for all eight conceptualizations. It is seen from this table that increasing the detail on the description of the PTA increases the posterior

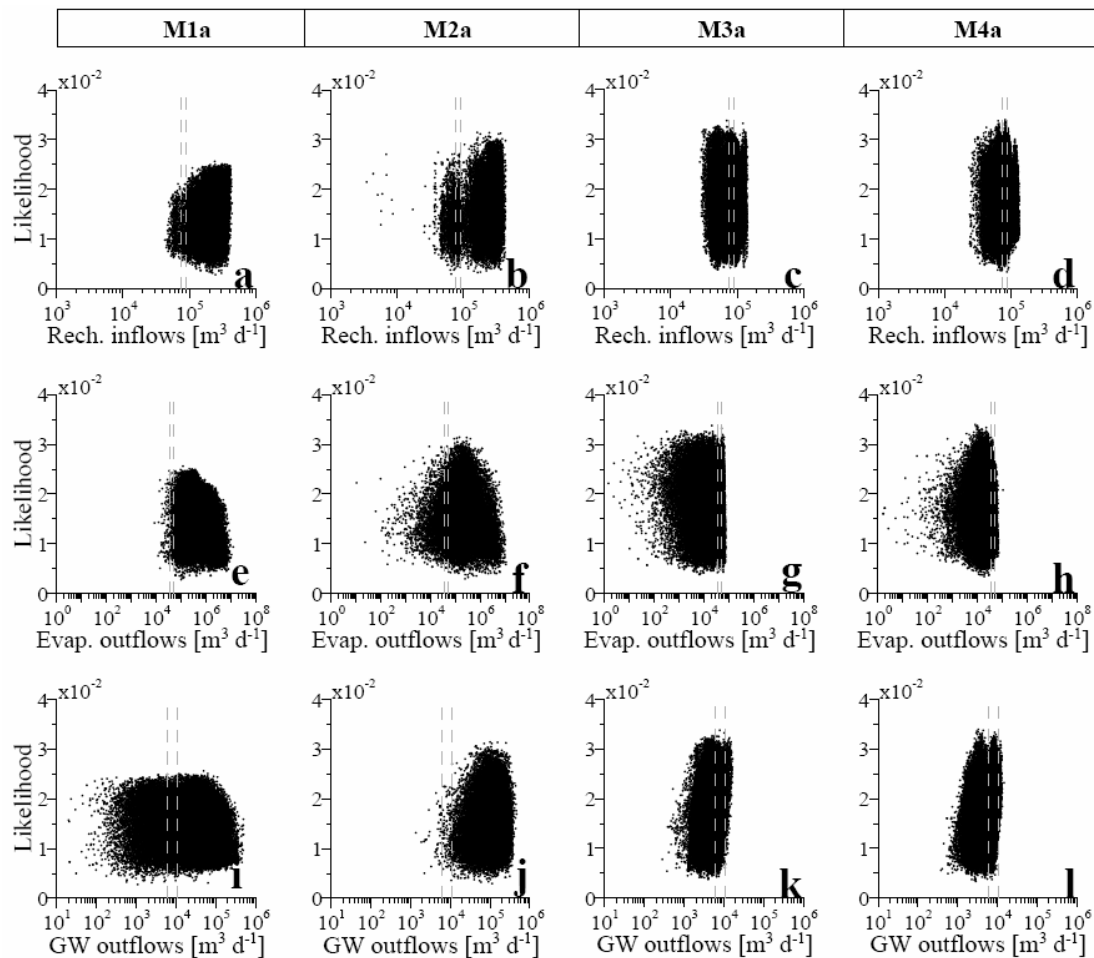


Fig. 6. Scatter plots of likelihood values for recharge inflows (a–d), evaporation outflows (e–h), and groundwater fluxes at the section defined between Cerro Gordo and the east boundary of the model (i–l) for models M1a, M2a, M3a, and M4a. Vertical dashed lines represent the range of estimated values from previous studies (DGA, 1988; JICA-DGA-PCI, 1995; DSM, 2002; DICTUC, 2005; Rojas and Dassargues, 2007; DICTUC, 2008).

model probabilities of the corresponding models. In this way, conceptual models using conditional realizations of the hydraulic conductivity field show higher posterior model probabilities. Although this might be attributed to the number of parameters used to describe the K -field, this is in agreement with the results obtained by Rojas et al. (2009a). In addition, when the alternative recharge mechanism is considered (models M1b, M2b, M3b and M4b), posterior model probabilities also increase for detailed descriptions of the PTA.

There seems to be no clear relationship between posterior model probabilities obtained for models including only recharge from the eastern sub-basins (a-version models) or models including both the recharge from the eastern sub-basins and the recharge due to deep fissures (b-version models). For example, models M1a and M3a show higher posterior weight compared to models M1b and M3b, respectively, whereas the opposite is observed for models M2a and M2b,

and models M4a and M4b. This indicates that the information content of the dataset D (42 observed heads) used to assess the alternative conceptual models does not allow a consistent discrimination between both recharge mechanisms. Therefore, other sources of information or conditioning data apart from head and hydraulic conductivity measurements should be considered. In this regard, Rojas et al. (2009a) have demonstrated that head measurements are limited in its ability to discriminate among conceptualizations included in M . This discrimination is drastically improved, however, by the inclusion of flow-related measurements and a sufficiently dense network of hydraulic conductivity measurements used for spatial conditioning. Unfortunately, for the problem at hand, such information was not available.

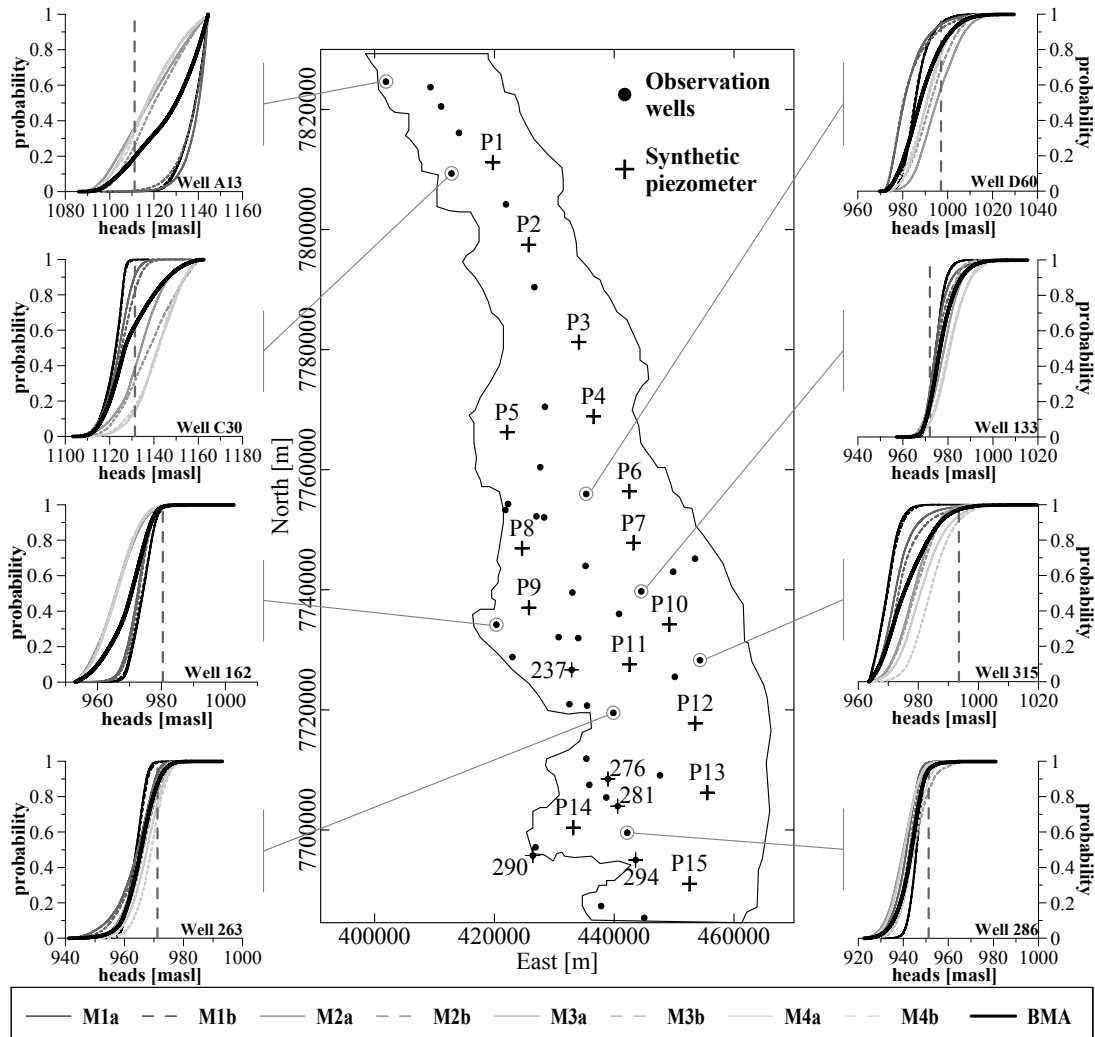


Fig. 7. Predictive distributions for a set of observation wells for the alternative conceptual models and the corresponding BMA predictive distribution. Vertical dashed lines represent observed head. Crosses represent location of synthetic piezometers.

5.4 Groundwater model predictions accounting for conceptual model uncertainties

5.4.1 Groundwater heads

Figure 7 shows a set of representative observation wells and the corresponding predictive distributions obtained for the alternative conceptualizations and the full BMA predictive distribution obtained from the GLUE-BMA methodology. Despite the fact heads at the observation wells were used as conditioning data, predictive distributions for the simulated groundwater heads varied significantly in spread, shape and central moment. The BMA predictive distribution accounts for these differences and, hence, represents a good compromise among all predictions obtained from the alternative conceptualizations. More important, deviations from this average prediction can be estimated for individual models,

hence, accounting for conceptual model uncertainty. There is a slight tendency to larger spreads of the predictive distributions for models M3b and M3a. The latter may be explained by the fact that model M3 describes the hydraulic conductivity fields using unconditional realizations following the spatial correlation structure shown in Fig. 4 solely. Thus, the level of spatial uncertainty is relatively high compared to the conditional case (model M4) or simpler models (M1 and M2).

Figure 7 also includes the location of 15 synthetic piezometers, which were included to explicitly assess the relevance of uncertainty arising from the definition of alternative conceptualizations at points not included in \mathcal{D} (heads + conductivity measurements). It is worth mentioning, however, that this set of synthetic piezometers were not intended to be used in a cross-validation approach. For such approach, an exact reproduction of heads (at observation

Table 4. Integrated model likelihoods ($p(\mathbf{D}|M_k)$), prior model probabilities ($p(M_k)$), and posterior model probabilities ($p(M_k|\mathbf{D})$) for the alternative conceptual models.

	Conceptual models								Total
	M1a	M1b	M2a	M2b	M3a	M3b	M4a	M4b	
$p(\mathbf{D} M_k)$	597.3	590.9	620.4	656.5	741.6	726.5	759.3	797.4	5498.84
$p(M_k)$	(1/8)	(1/8)	(1/8)	(1/8)	(1/8)	(1/8)	(1/8)	(1/8)	1.0
$p(M_k \mathbf{D})$	0.1088	0.1076	0.1130	0.1196	0.1351	0.1323	0.1382	0.1452	1.0

wells or synthetic piezometers) would have been necessary. Since GLUE-BMA allows for a departure from observed data by defining tolerable limits through the rejection criterion, such cross-validation approach is not feasible in the context of this work. Although not shown here, predictive distributions for these synthetic piezometers varied substantially in shape, central moment and spread due to the definition of the alternative conceptualizations. For these synthetic piezometers the contribution of conceptual model uncertainty was relatively important as individual model predictions at these locations varied substantially from the average prediction (see e.g. Fig. 10b and Tables 5 and 6). The latter indicates that conceptual model uncertainty is more significant for spatial points not included as conditioning data. It is worth emphasizing that we were interested in the ratio between-model variance to total variance at these synthetic piezometers, which ultimately assess the relevance (contribution) of conceptual model uncertainty. This ratio can potentially guide data collection campaigns aiming at reducing this source of uncertainty.

Spatial head distributions for the eight alternative conceptualizations and the BMA prediction are shown in Fig. 8. In general, models M1 and M2 (Fig. 8a–d) show a consistent spatial head distribution with smooth contour lines. Two weak points of these results, however, are the poor representation of heads at the northern and eastern parts of the modelled domain. Representing the K -field as an RSF improves the spatial distribution of heads and models M3 and M4 (Fig. 8e–h) show a better representation of the northern and eastern parts. It is worth noticing that model M4b (Fig. 8h), which shows the highest posterior model probability (0.1452), shows the best representation of the spatial head distribution compared to the observed heads. A weak point of the results from models including an RSF representation is that equipotential head lines are considerably less smooth and more irregular than the observed spatial head distribution. A good compromise between the eight spatial head distributions is obtained with the BMA prediction (Fig. 8i). The latter combines the consistency and smoothness of the results of models M1 and M2, and the improved spatial head distributions obtained from models M3 and M4. It is worth noting, however, that the BMA prediction also shows a poor representation of the spatial head distribution in the north-

ern area and to a lesser extent in the eastern part. This is due to the fact that model weights used for multi-model aggregation were rather similar, with a cumulative weight for models M1 and M2 of ca. 0.45. The latter indicates that no single model is preferred over the others on the basis of the evidence provided by the dataset \mathbf{D} . As discussed earlier, this problem could be solved by assigning non-uniform prior model distributions reflecting the analyst's prior perception about the plausibility of the alternative conceptualizations or by collecting more data on key properties of the alternative conceptualizations.

5.4.2 Groundwater flow components

In general, predictive distributions for the groundwater flow components differed significantly among models indicating an important contribution of conceptual model uncertainty to the predictive uncertainty (Fig. 9). Models M3 and M4 showed a better predictive coverage of the ranges estimated in previous studies compared to models M1 and M2. For transpiration outflows, however, the lower limit of the range defined by previous estimations was defined rather at the tail of all individual and the BMA predictions. This might suggest that higher transpiration outflows were observed for year 1960 compared to the lower limit estimated by DICTUC (2008) for transpiration outflows. It is worth pointing out, however, that the estimation made in DICTUC (2008) (ca. $1728 \text{ m}^3 \text{ d}^{-1}$) is rather conservative compared to transpiration outflows ($18\,144 \text{ m}^3 \text{ d}^{-1}$) used in previous studies (DGA, 1988; JICA-DGA-PCI, 1995; Rojas and Dassargues, 2007).

For the groundwater outflows to La Noria aquifer (Fig. 9e), all conceptualizations showed a similar (in spread and central moment) predictive distribution. In addition, the most reliable estimation available for these outflows ($1555 \text{ m}^3 \text{ d}^{-1}$) (DICTUC, 2008) was located at the lower tail of the predictive distributions, indicating that it is likely that the outflows from PTA to La Noria aquifer might be higher than the current estimation. It is important to note that the estimation made by DICTUC (2008) corresponds to present situation where the groundwater elevations for the PTA are relatively deeper compared to the situation in 1960. Thus, it is likely that for year 1960 the groundwater outflows from PTA to

Table 5. Summary of observed heads, BMA prediction and predictive variances for groundwater heads at observation wells (see Fig. 1b).

	Observation well nr.												
	162	237	263	276	281	286	290	294	C-30	315	D-60	A-13	133
Observed [m]	980.4	967.8	971.2	957.5	954.9	951.3	951.9	924.5	1131.4	993.5	997.1	1111.2	972.0
BMA prediction	968.1	966.4	964.7	950.6	945.7	942.1	937.5	932.8	1130.1	975.8	987.6	1125.3	976.5
± std. dev [m]	±6.5	±4.9	±5.5	±5.6	±6.0	±6.0	±5.1	±7.9	±11.6	±7.8	±9.3	±14.5	±6.4
Predictive variance [m ²]	41.8	23.6	30.0	31.9	36.5	36.2	26.5	62.8	134.8	60.3	89.6	211.7	41.2

Table 6. Summary of BMA prediction and predictive variances of groundwater heads at synthetic piezometers (see Fig. 1b).

	Synthetic piezometers											
	P1	P2	P3	P4	P7	P10	P11	P12	P13	P14	P15	
BMA prediction	1131.0	1101.2	1043.3	1009.8	981.1	972.3	973.2	965.8	943.3	942.3	929.7	
± std. dev. [m]	±13.8	±18.6	±17.3	±14.5	±7.5	±5.5	±7.8	±8.7	±9.4	±10.1	±12.1	
Predictive variance [m ²]	191.0	345.9	299.7	209.7	54.6	30.7	60.9	76.5	88.6	102.0	146.9	

La Noria were somewhat higher due to possibly steeper hydraulic gradients in the section defining the hydrogeological connection between PTA and La Noria aquifers.

5.5 Contribution to predictive variance

The predictive variance for groundwater heads at observation points, for the synthetic piezometers, and for the groundwater flow components are shown in Tables 5, 6, and 7, respectively. Estimations of groundwater heads at the north area of the modelled domain showed to be the most uncertain, presenting the highest variances for observation wells C-30 and A-13 and piezometers P1, P2 and P3 (see Fig. 10). This is due to the effects of groundwater recharge inflows from two major eastern sub-basins located in that area (Aroma and Tarapaca), which on average accounts for more than 60% of the total recharge to the PTA. Groundwater recharge inflows derived from these sub-basins considerably affect the uncertainty in the estimation of the groundwater heads in the northern area. The latter is in full agreement with the results obtained by Rojas and Dassargues (2007). Also at the southern area of the modeled domain relatively high variances were observed for observation well 294 and piezometers P13, P14 and P15, indicating high uncertainty in the estimation of the heads in this area. This may partially be explained by the presence of the recharge front originated from the Chacarilla sub-basin (see Fig. 1), which on average accounts for 20% of the total recharge to the PTA, and also due to the proximity of the south boundary condition.

For observation wells the contribution of variance derived from conceptual model uncertainty to the predictive variance varied between 10% and 45%, whereas for piezometers

ranged between 6% and 64% (Fig. 10). In the north area, the definition of alternative conceptual models showed a significant impact on the uncertainty estimations (see observation wells C-30 and A-13 in Fig. 10a). It is worth noticing, however, that for observation wells 162 and 315 conceptual model uncertainty also significantly contributed to the predictive variance of head estimations. Observation well 315 is located in the distal part of the alluvial fan formed in the Chacarilla sector, hence, it is influenced by the groundwater recharge inflows originating from this sub-basin. Therefore, different ways to represent these recharge inflows will have an impact on the uncertainty estimation for the head at this well. Observation well 162, on the contrary, is located at the western side of the modeled domain in the Salar de Pintados area, near the discharge points connecting PTA and La Noria aquifers (see Figs. 1 and 7). This proximity to these discharge points could only explain the contribution of within-models variance to the predictive variance in the estimation of the head at this well, since by definition these discharging points are a common element for all eight alternative conceptualizations. In addition, if recharge inflows originating from the Chacarilla sub-basin were influencing the estimation of conceptual model uncertainty in well 162, a similar influence should also be noticed for the wells located in between wells 315 and 162. However, this is not observed. Therefore, the only probable explanation for this large contribution of conceptual model uncertainty to the predictive uncertainty in well 162 is the differences in the representation of the hydraulic conductivity field among all eight alternative conceptualizations. This suggests that for areas not directly affected by recharge fronts, the representation of

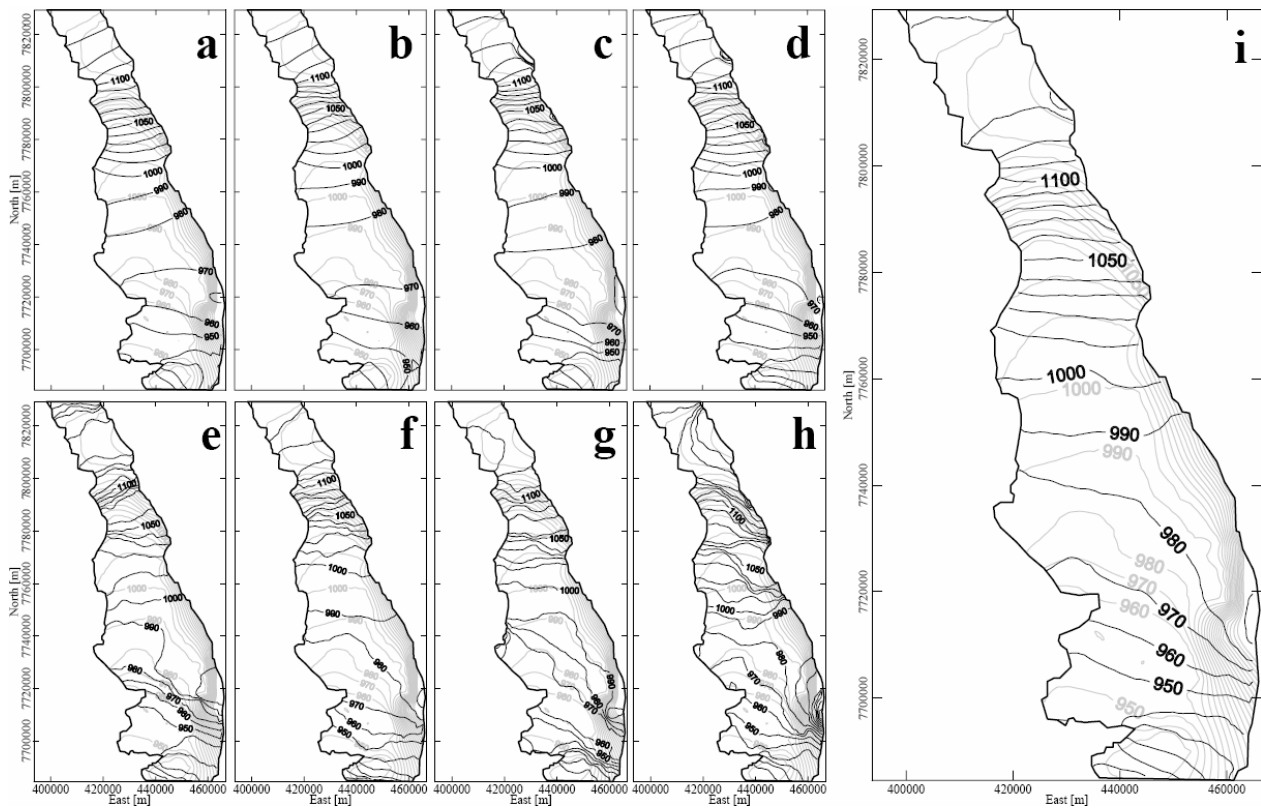


Fig. 8. Most likely predicted spatial head distribution for models (a) M1a, (b) M1b, (c) M2a, (d) M2b, (e) M3a, (f) M3b, (g) M4a, (h) M4b and (i) BMA.

Table 7. Summary of predictive variances for estimations of groundwater flow components. Values expressed in $(\text{m}^3 \text{d}^{-1})^2$.

GW flow component	Predictive variance
Recharge inflows	1.93×10^{10}
Evaporation outflows	4.39×10^{11}
Transpiration outflows	2.80×10^9
Recharge Chacarilla sub-basin	3.26×10^8
Discharge to La Noria aquifer	5.41×10^8
Recharge from deep fissures	7.47×10^9
GW outflows at Cerro Gordo	4.23×10^9

the hydraulic conductivity field dominates the contribution of conceptual model uncertainty to predictive variance in the estimation of groundwater heads.

Results from Fig. 10 can be thought of as a split-sample test using multiple conceptual models, where forty-two data values are used for obtaining the model weights (posterior model probabilities) for multi-model aggregation and fifteen data values are used as a pseudo-validation-sample (see Fig. 7). Results show that for the pseudo-validation sample (synthetic piezometers) the predictive variances are sig-

nificantly higher compared to predictive variances obtained using the forty-two observation wells. This indicates the relevance of conceptual model uncertainty for spatial data not included as conditioning points (pseudo-validation-sample).

Figure 11 shows the percentage contribution of conceptual model uncertainty to the predictive variance for the groundwater flow components of interest. In general, uncertainty arising from the alternative conceptualizations was significantly more important for the recharge inflows accounting for 76% and 79% of the predictive variance for the estimation of recharge originating from the Chacarilla sub-basin and for the effective recharge inflows from all sub-basins, respectively. For the recharge inflows, the most uncertain (individual) estimations were linked to simpler models M1 and M2, whereas the most accurate estimations were associated to models M3 and M4 (see Fig. 9). This shows the relevance of describing the hydraulic conductivity field following a RSF to obtain more confident recharge estimations. If simpler models are used, it is likely that recharge estimations will be highly uncertain, therefore, more effort should be invested in describing the recharge mechanisms when working with simpler conceptualizations to approximate the PTA.

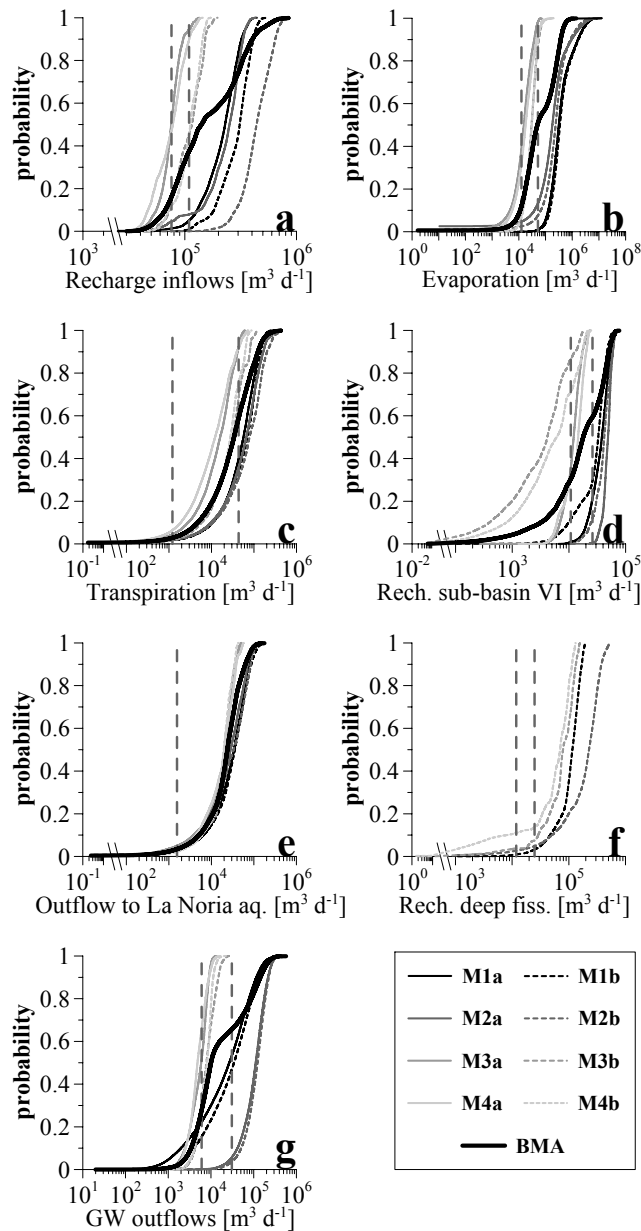


Fig. 9. Predictive cumulative distributions for groundwater flow components: (a) recharge inflows, (b) evaporation outflows, (c) transpiration outflows, (d) recharge Chacarilla sub-basin, (e) discharge to La Noria aquifer, (f) recharge due to faults and deep fissures, and (g) groundwater outflows in the eastern section of Cerro Gordo. Vertical dashed lines represent the range of estimated values from previous studies (DGA, 1988; JICA-DGA-PCI, 1995; DSM, 2002; DICTUC, 2005; Rojas and Dassargues, 2007; DICTUC, 2008).

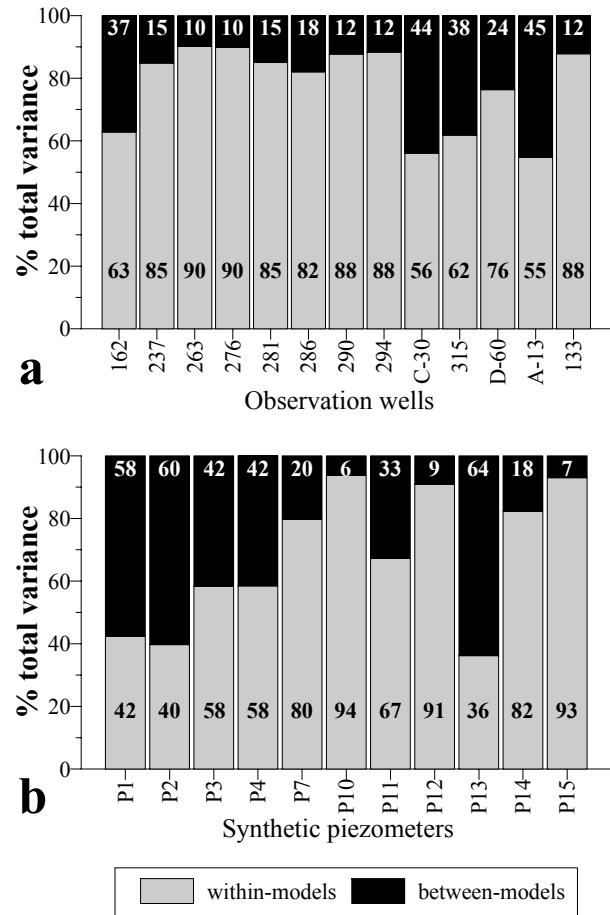


Fig. 10. Predictive variance obtained using Eq. (4) expressed as a percentage contribution from within-models and between-models components for: (a) observation wells and (b) synthetic piezometers depicted in Fig. 1b.

Estimation of groundwater outflows defined at the southern section of Cerro Gordo also showed a large contribution of conceptual model uncertainty (55%). For this flow component, the largest individual variances are associated to models M2 and M1, respectively. Neglecting this contribution of between-models variances to the predictive uncertainty may have serious implications on groundwater management of the study area since these groundwater fluxes are considered the main recharge inflows to the southern aquifer of Salar Viejo.

The contribution of conceptual model uncertainty to the predictive variance for the transpiration outflows resulted in the order of 30% with the largest individual variance obtained for model M2. For evaporation outflows and discharges to La Noria aquifer, contributions of between-models variance were 19% and 16%, respectively. It is worth noticing that even for the case of the outflows to La Noria, where predictive distributions were rather similar between the alternative conceptualizations, a moderate contribution of conceptual model uncertainty was observed.

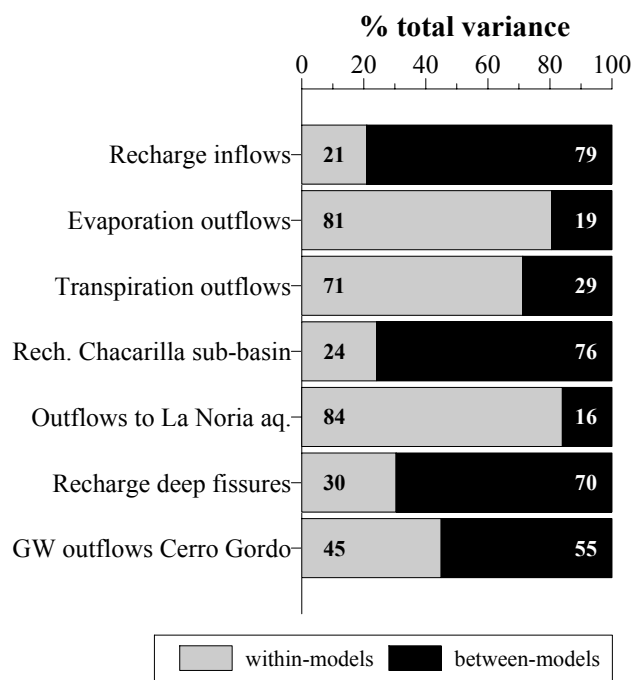


Fig. 11. Contribution of within-models and between-models variances (expressed as a percentage of the predictive variance) for main groundwater flow components of the PTA.

Figure 11 illustrates the importance of conceptual model uncertainty when making extrapolations beyond the dataset used for calibration. The main dataset used for estimation of posterior model probabilities were groundwater heads whereas model predictions were obtained for flow components. These results show the relevance of considering alternative conceptual models for predictions of variables not used as calibration targets and are in full agreement with the findings of Harrar et al. (2003), Troldborg et al. (2007) and Seifert et al. (2008).

6 Summary and conclusions

In this work we assessed the uncertainty in groundwater flow modelling of the regional aquifer Pampa del Tamarugal located in northern Chile using a multi-model methodology aimed at explicitly accounting for conceptual model uncertainty. We proposed an ensemble (M) of eight alternative conceptualizations covering all major features of groundwater flow models previously developed, ranging from a simple two-layer representation to models fully accounting for the spatial heterogeneity of the hydraulic conductivity field. We further developed models representing the heterogeneity of K -field by conditioning the realizations to available hydraulic conductivity measurements. We included two recharge mechanisms, which have been source of debate for several years to account for uncertainties in the recharge in-

flows to the groundwater system. For each member of the ensemble M , integrated model likelihoods and posterior model probabilities were derived, which were then used to obtain multi-model predictions that explicitly account for conceptual model uncertainty.

By definition, results of the GLUE-BMA methodology are conditional on the proposed ensemble M of alternative conceptual models, hence, the “quality” of the GLUE-BMA predictions is linked to the comprehensiveness of the ensemble M . That is, if the members of M cover a suitable range of potential conceptualizations while ensuring that they are different enough to consider them mutually exclusive, the quality of the uncertainty assessment will be improved. On the contrary, if models are rather similar (under-sampling) or tend to reflect fairly similar processes or features (biased), the “quality” of the uncertainty assessment using the GLUE-BMA method will be weakened.

We acknowledge that other model structures could be included in the analysis, for example, other spatial distributions (zonations) to describe the hydraulic conductivity field, a full two-layer description accounting for small-scale heterogeneities or even a larger modelled domain to explicitly connect the PTA with aquifer systems located to the south or west (e.g. Salar Viejo, Salar de Llamara and La Noria). First, any spatial distribution (zonation) of the hydraulic conductivity field obtained through a scientifically sound method (e.g. model calibration) is considered a valid representation of the K -field. Hence, alternative zonations are equally likely and valid in the frame of this uncertainty assessment. For pragmatic reasons, a spatial zonation available from Rojas and Dassargues (2007) was selected. Second, a two-layer description accounting for small-scale heterogeneities would be suitable if the dataset used to characterize separately both units (Q3 and Q4) of the PTA was comprehensive. Given the dimensions of the aquifer system, the current level of information is scarce and the number of available hard data on key parameters is at its limit to perform a meaningful geostatistical analysis. Thus, it seems more conservative to combine both units (Q3 and Q4) in one hydrostratigraphic unit and model PTA as a regional two-dimensional system. If more hard data on key parameters were (independently) collected for both units, a two-layer conceptualization fully accounting for the heterogeneities on both units would be worth exploring. Third, including a larger model domain would imply to increase the dataset D to assess the model performance as more information would be available from southern or western aquifer systems. To correctly assess the conceptual model uncertainty, however, a common dataset D to all conceptualizations should be used since models including more data to assess the model performance would be implicitly accounting for the worth of the additional data, hence, masking the actual weight of the conceptualization.

We recognize that the GLUE-BMA approach does not explicitly penalize (through the likelihood function) models with a high number of parameters in order to comply with

the principle of parsimony. This, however, can be achieved by defining non-uniform prior model probabilities, which should reflect the analyst's prior perception about the plausibility of the alternative conceptualizations. How efficiently define these priors will be the subject of future research.

By mimicking actual conceptual models developed in the last 20 years for the PTA we demonstrated that conceptual model uncertainty is a relevant source of predictive uncertainty in this area. This is particularly important for the PTA where human pressure for water resources is considerably high and uncertainty due to climatic conditions is relatively important. As a consequence, it is recommended to acknowledge this source of uncertainty to provide a more robust and sustainable management of the groundwater reserves of PTA.

The main findings of this work can be summarized as follows:

1. Considering a more detailed description of the heterogeneity of the hydraulic conductivity fields reduced the spread of the likelihood response surfaces, hence, decreasing the uncertainty in the estimations of parameters and state variables of concern for the PTA. In addition, accounting for the heterogeneity and conditioning the hydraulic conductivity field on available conductivity measurements increased the posterior model probabilities.
2. The GLUE-BMA predictive distributions encompassed estimations made in previous studies, thus, showing consistence with previous knowledge about the groundwater system.
3. Models M1 and M2 failed to adequately characterize the range of estimations from previous studies for transpiration outflows, groundwater recharge from deep fissures and the outflows from PTA to La Noria aquifer.
4. A set of 42 head observations did not allow a clear discrimination between the two recharge mechanisms considered in this study. To further differentiate about the validity of both recharge mechanisms other sources of information or conditioning data should be considered. A better characterization of both Q3 and Q4 units to independently condition the K -fields will likely result in a better discrimination between the recharge mechanisms.
5. There seems to be an apparent spatial relationship between the level of uncertainty in the estimations of groundwater heads and areas directly affected by recharge fronts.
6. For areas not affected by the recharge fronts, conceptual model uncertainty seems to be driven by the alternative representations of the hydraulic conductivity field among models.
7. The relevance of conceptual model uncertainty is more important for prediction of variables not used as calibration targets, i.e. extrapolations beyond the dataset used for calibration. These results are in full agreement with the findings of Harrar et al. (2003), Troldborg et al. (2007) and Seifert et al. (2008).
8. Contribution of conceptual model uncertainty varied between 16% and 79% of the predictive variance for the groundwater flow components whereas for the estimations of heads at the observation wells and at the synthetic piezometers varied between 7% and 64% of the predictive variance.

Finally, a better quantitative hydrogeological description of both units Q3 and Q4 to improve the geostatistical modelling, a validation (quantification) of the recharge mechanisms involved in the groundwater dynamics of the PTA, and the extension of the modelled domain to the south in order to include flow-related observations are interesting future research alternatives.

Acknowledgements. The first author thanks the Katholieke Universiteit Leuven (K.U. Leuven) for providing financial support in the framework of PhD IRO-scholarships. We also wish to thank A. M. Gangas of the Dirección General de Aguas (DGA) Chile, for helpful comments and insightful discussions about current studies carried out on the Pampa del Tamarugal Aquifer. This research was conducted utilizing high performance computational resources provided by K.U. Leuven through the VIC Cluster, <http://ludit.kuleuven.be/hpc>.

Edited by: A. Guadagnini

References

- Aravena, R.: Isotope hydrology and geochemistry of northern Chile groundwaters, *Bull. IFEA*, 24(3), 495–503, 1995.
- Aravena, R., Peña, H., Grilli, A., Suzuki, O., and Mordeckai, M.: Evolución isotópica de las lluvias y origen de las masas de aire en el altiplano chileno, in: *Isotope Hydrology Investigations In Latin America*, IAEA, TECDOC-502, Vienna, Austria, 14 pp., 1989.
- Aravena, R., Suzuki, O., Peña, H., Pollastri, A., Fuenzalida, H., and Grilli, A.: Isotopic composition and origin of the precipitation in northern Chile, *Appl. Geochem.*, 14(4), 411–422, doi:10.1016/S0883-2927(98)00067-5, 1999.
- Beven, K.: A manifesto for the equifinality thesis, *J. Hydrol.*, 320(1–2), 18–36, doi:10.1016/j.jhydrol.2005.07.007, 2006.
- Beven, K. and Binley, A.: The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Process.*, 6(5), 279–283, doi:10.1002/hyp.3360060305, 1992.
- Beven, K. and Freer, J.: Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, *J. Hydrol.*, 249(1–4), 11–29, doi:10.1016/S0022-1694(01)00421-8, 2001.

- Binley, A. and Beven, K.: Vadose zone flow model uncertainty as conditioned on geophysical data, *Ground Water*, 41(2), 119–127, doi:10.1111/j.1745-6584.2003.tb02576.x, 2003.
- Bredehoeft, J.: From models to performance assessment: The conceptualization problem, *Ground Water*, 41(5), 571–577, doi:10.1111/j.1745-6584.2003.tb02395.x, 2003.
- Bredehoeft, J.: The conceptualization model problem-surprise, *Hydrogeol. J.*, 13(1), 37–46, doi:10.1007/s10040-004-0430-5, 2005.
- Chib, S. and Greenberg, E.: Understanding the Metropolis-Hastings algorithm, *Am. Stat.*, 49(4), 327–335, <http://www.jstor.org/stable/2684568>, 1995.
- Chilès, J. and Delfiner, P.: *Geostatistics: Modeling Spatial Uncertainty*, 1st edn., John Wiley and Sons, New York, USA, 720 pp., 1999.
- DGA: Modelo de simulación hidrogeológica de la Pampa del Tamarugal, Dirección General de Aguas, Santiago, Chile, 98 pp., 1988.
- DGA: Determinación de la disponibilidad de recursos hídricos para constituir nuevos derechos de aprovechamiento de aguas subterráneas en el sector del acuífero de la Pampa del Tamarugal, S.D.T. Nr. 68, Dirección General de Aguas, Santiago, Chile, 27 pp., 1996.
- DICTUC: Evaluación del impacto del proyecto de extracción de aguas subterráneas en la zona sur de la Pampa del Tamarugal, Tech. rep., Dirección de Investigaciones Científicas y Tecnológicas Universidad Católica de Chile, Santiago, Chile, 1995.
- DICTUC: Análisis técnico de solicitud de derechos de aprovechamiento de ACF Minera. Acuífero Pampa del Tamarugal, Dirección de Investigaciones Científicas y Tecnológicas Universidad Católica de Chile, Santiago, Chile, 107 pp., 2005.
- DICTUC: EIA Proyecto Pampa Hermosa. Anexo IX.1 Geología del acuífero de la Pampa del Tamarugal, Dirección de Investigaciones Científicas y Tecnológicas Universidad Católica de Chile, Santiago, Chile, <https://www.e-seia.cl/expediente/expedientes.php?modo=ficha&id.expediente=3083858&idExpediente=3083858>, 99 pp., 2007.
- DICTUC: EIA Proyecto Pampa Hermosa. Anexo VIII.2 Hidrogeología, Dirección de Investigaciones Científicas y Tecnológicas Universidad Católica de Chile, Santiago, Chile, <https://www.e-seia.cl/expediente/expedientes.php?modo=ficha&id.expediente=3083858&idExpediente=3083858>, 169 pp., 2008.
- Digert, F., Hoke, G., Jordan, T., and Isaaks, B.: Subsurface stratigraphy of the neogene Pampa del Tamarugal basin, northern Chile, in: *X Congreso Geológico Chileno*, Concepción, Chile, 8 pp., 2003.
- Dingman, R. and Galli, C.: Geology and groundwater resources of the Pica area, Tarapaca Province, Chile, Tech. rep., US Geological Survey, Virginia, USA, 125 pp., 1965.
- Draper, D.: Assessment and propagation of model uncertainty, *J. Roy. Stat. Soc. B*, 57(1), 45–97, <http://www.jstor.org/stable/2346087>, 1995.
- DSM: Estación Cerro Gordo. Modelo de flujo numérico de aguas subterráneas, DSM Minera S.A., Santiago, Chile, 70 pp., 2002.
- FAO: Role of forestry in combating desertification: Proceedings of the FAO Expert Consultation on the Role of Forestry in Combating Desertification held in Saltillo, Mexico, 24–28 June 1985, Tech. rep., FAO, Rome, Italy, <http://www.fao.org/docrep/t0115e/T0115E00.htm>, 1989.
- Feyen, L., Beven, K., De Smedt, F., and Freer, J.: Stochastic capture zone delineation within the GLUE-methodology: Conditioning on head observations, *Water Resour. Res.*, 37(3), 625–638, <http://www.agu.org/pubs/crossref/2001/2000WR900351.shtml>, 2001.
- Fritz, P., Suzuki, O., Silva, C., and Salati, E.: Isotope hydrology of groundwaters in the Pampa del Tamarugal, Chile, *J. Hydrol.*, 53(1–2), 161–184, doi:10.1016/0022-1694(81)90043-3, 1981.
- Gelman, A., Carlin, J., Stern, H., and Rubin, D.: *Bayesian Data Analysis*, 2nd edn., Chapman and Hall/CRC, New York, USA, 696 pp., 2004.
- Ghosh, J., Delampady, M., and Samanta, T.: *An Introduction to Bayesian Analysis – Theory and Methods*, 1st edn., Springer-Verlag, New York, USA, 352 pp., 2006.
- Gilks, W., Richardson, S., and Spiegelhalter, D.: *Markov Chain Monte Carlo in Practice*, 1st edn., Chapman and Hall/CRC, Boca Raton, Florida, USA, 512 pp., 1995.
- Harbaugh, A., Banta, E., Hill, M., and McDonald, M.: *MODFLOW-2000 US Geological Survey modular ground-water model-user guide to modularization concepts and the ground-water flow process*, Open file rep., 00–92, United States Geological Survey, Reston, Virginia, USA, 2000.
- Harrar, W., Sonnenberg, T., and Henriksen, H.: Capture zone, travel time, and solute transport predictions using inverse modelling and different geological models, *Hydrogeol. J.*, 11(5), 536–548, doi:10.1007/s10040-003-0276-2, 2003.
- Hastings, W.: Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, 57(1), 97–109, doi:10.1093/biomet/57.1.97, 1970.
- Hoeting, J., Madigan, D., Raftery, A., and Volinsky, C.: Bayesian model averaging: A tutorial, *Stat. Sci.*, 14(4), 382–401, <http://www.jstor.org/stable/2676803>, 1999.
- Højberg, A. and Refsgaard, J.: Model uncertainty-parameter uncertainty versus conceptual models, *Water Sci. Technol.*, 52(6), 177–186, <http://www.iwaponline.com/wst/05206/wst052060177.htm>, 2005.
- Houston, J.: Groundwater recharge through an alluvial fan in the Atacama Desert, northern Chile: mechanisms, magnitudes and causes, *Hydrol. Process.*, 16(15), 3019–3035, doi:10.1002/hyp.1086, 2002.
- Houston, J.: The great Atacama flood of 2001 and its implications for Andean hydrology, *Hydrol. Process.*, 20(3), 591–610, doi:10.1002/hyp.5926, 2006.
- Ijiri, Y., Saegusa, H., Sawada, A., Ono, M., Watanabe, K., Karasaki, K., Doughty, C., Shimo, M., and Fumimura, K.: Evaluation of uncertainties originating from the different modeling approaches applied to analyze regional groundwater flow in the Tono area of Japan, *J. Contam. Hydrol.*, 103(3–4), 168–181, doi:10.1016/j.jconhyd.2008.10.010, 2009.
- JICA-DGA-PCI: The study on the development of water resources in northern Chile, Supporting report B: Geology and groundwater, Japanese International Cooperation Agency, Dirección General de Aguas, and Pacific Consultants International, Santiago, Chile, 216 pp., 1995.
- Kass, R. and Raftery, A.: Bayes factors, *J. Am. Stat. Assoc.*, 90(430), 773–795, <http://www.jstor.org/stable/2291091>, 1995.
- Magaritz, M., Aravena, R., Peña, H., Suzuki, O., and Grilli, A.: Source of groundwaters in the deserts of northern Chile:

- Evidence for deep circulation of groundwaters from the Andes, *Ground Water*, 28(4), 513–517, doi:10.1111/j.1745-6584.1990.tb01706.x, 1990.
- Makowski, D., Wallach, D., and Tremblay, M.: Using a Bayesian approach to parameter estimation; comparison of the GLUE and MCMC methods, *Agronomie*, 22(2), 191–203, doi:10.1051/agro:2002007, 2002.
- Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E.: Equation of state calculations by fast computing machines, *J. Chem. Phys.*, 21(6), 1087–1092, doi:10.1063/1.1699114, 1953.
- Meyer, P., Ye, M., Rockhold, M., Neuman, S., and Cantrell, K.: Combined estimation of hydrogeologic conceptual model parameter and scenario uncertainty with application to uranium transport at the Hanford Site 300 area, Report NUREG/CR-6940 PNNL-16396, US Nuclear Regulatory Commission, Washington, US, 125 pp., 2007.
- Neuman, S.: Maximum likelihood Bayesian averaging of uncertain model predictions, *Stoch. Env. Res. Risk A*, 17(5), 291–305, doi:10.1007/s00477-003-0151-7, 2003.
- Neuman, S. and Wierenga, P.: A comprehensive strategy of hydrogeologic modelling and uncertainty analysis for nuclear facilities and sites, Report NUREG/CR-6805, US Nuclear Regulatory Commission, Washington, USA, 309 pp., 2003.
- Poeter, E. and Anderson, D.: Multimodel ranking and inference in ground water modelling, *Ground Water*, 43(4), 597–605, doi:10.1111/j.1745-6584.2005.0061.x, 2005.
- Refsgaard, J., Van der Sluijs, J., Brown, J., and Van der Keur, P.: A framework for dealing with uncertainty due to model structure error, *Adv. Water Resour.*, 29(11), 1586–1597, doi:10.1016/j.advwatres.2005.11.013, 2006.
- Refsgaard, J., Van der Sluijs, J., Højberg, A., and Vanrolleghem, P.: Uncertainty in the environmental modelling process—A framework and guidance, *Environ. Modell. Softw.*, 22(11), 1543–1556, doi:10.1016/j.envsoft.2007.02.004, 2007.
- Risacher, F., Alonso, H., and Salazar, C.: The origin of brines and salts in Chilean salar: A hydrochemical review, *Earth Sci. Rev.*, 63(3–4), 249–293, doi:10.1016/S0012-8252(03)00037-0, 2003.
- Rojas, R. and Dassargues, A.: Groundwater flow modelling of the regional aquifer of the Pampa del Tamarugal, northern Chile, *Hydrogeol. J.*, 15(3), 537–551, doi:10.1007/s10040-006-0084-6, 2007.
- Rojas, R., Feyen, L., and Dassargues, A.: Conceptual model uncertainty in groundwater modelling: Combining generalized likelihood uncertainty estimation and Bayesian model averaging, *Water Resour. Res.*, 44, W12418, doi:10.1029/2008WR006908, 2008.
- Rojas, R., Feyen, L., Batelaan, O., and Dassargues, A.: On the value of conditioning data to reduce conceptual model uncertainty in groundwater modelling, *Water Resour. Res.*, under review, 2009a.
- Rojas, R., Feyen, L., and Dassargues, A.: Sensitivity analysis of prior model probabilities and the value of prior knowledge in the assessment of conceptual model uncertainty in groundwater modelling, *Hydrol. Process.*, 23(8), 1131–1146, doi:10.1002/hyp.7231, 2009b.
- Rojas, R., Kahunde, S., Peeters, L., Batelaan, O., and Dassargues, A.: Application of a multimodel approach to account for conceptual model and scenario uncertainties in groundwater modelling, *J. Hydrol.*, under review, 2009c.
- Salazar, C., Rojas, L., and Pollastri, A.: Evaluación de recursos hídricos en el sector de Pica hoyo de la Pampa del Tamarugal I Región-Chile, in: VI Jornadas del CONAFI Chile, Santiago, Chile, 16 pp., 1999.
- Seifert, D., Sonnenberg, T., Scharling, P., and Hinsby, K.: Use of alternative conceptual models to assess the impact of a buried valley on groundwater vulnerability, *Hydrogeol. J.*, 16(4), 659–674, doi:10.1007/s10040-007-0252-3, 2008.
- Sorensen, D. and Gianola, D.: Likelihood, Bayesian, and MCMC methods in quantitative genetics, Vol. I, 1st edn., Springer-Verlag, New York, USA, 760 pp., 2002.
- Suzuki, O. and Aravena, R.: Hidrología isotópica y el recurso agua del sector Esmeralda-Pica-Matilla, *Nucleotécnica*, 4(8), 41–51, 1985.
- Troldborg, L., Refsgaard, J., Jensen, K., and Engesgaard, P.: The importance of alternative conceptual models for simulation of concentrations in a multi-aquifer system, *Hydrogeol. J.*, 15(5), 843–860, doi:10.1007/s10040-007-0192-y, 2007.
- Wasserman, L.: Bayesian model selection and model averaging, *J. Math. Psychol.*, 44(1), 92–107, doi:10.1006/jmps.1999.1278, 2000.
- Ye, M., Meyer, P., and Neuman, S.: On model selection criteria in multimodel analysis, *Water Resour. Res.*, 44, W03428, doi:10.1029/2008WR006803, 2008a.
- Ye, M., Neuman, S., and Meyer, P.: Maximum likelihood Bayesian averaging of spatial variability models in unsaturated fractured tuff, *Water Resour. Res.*, 40, W05113, doi:10.1029/2003WR002557, 2004.
- Ye, M., Pohlmann, K., and Chapman, J.: Expert elicitation of recharge model probabilities for the Death Valley regional flow system, *J. Hydrol.*, 354(1–4), 102–115, doi:10.1016/j.jhydrol.2008.03.001, 2008b.