
Stochastic simulation of time-series models combined with geostatistics to predict water-table scenarios in a Guarani Aquifer System outcrop area, Brazil

Rodrigo L. Manzione · Edson Wendland ·
Diego H. Tanikawa

Abstract Stochastic methods based on time-series modeling combined with geostatistics can be useful tools to describe the variability of water-table levels in time and space and to account for uncertainty. Monitoring water-level networks can give information about the dynamic of the aquifer domain in both dimensions. Time-series modeling is an elegant way to treat monitoring data without the complexity of physical mechanistic models. Time-series model predictions can be interpolated spatially, with the spatial differences in water-table dynamics determined by the spatial variation in the system properties and the temporal variation driven by the dynamics of the inputs into the system. An integration of stochastic methods is presented, based on time-series modeling and geostatistics as a framework to predict water levels for decision making in groundwater management and land-use planning. The methodology is applied in a case study in a Guarani Aquifer System (GAS) outcrop area located in the southeastern part of Brazil. Communication of results in a clear and understandable form, via simulated scenarios, is discussed as an alternative, when translating scientific knowledge into applications of stochastic

hydrogeology in large aquifers with limited monitoring network coverage like the GAS.

Keywords Groundwater monitoring · Geostatistics · Statistical modeling · Brazil

Introduction

The application of stochastic approaches to hydrogeology is emerging from some parts of academia and starting to appear more frequently in discussion meetings of water commissions and regulatory bodies involving decision makers, policy makers and stakeholders. In a changing world, professionals are becoming more open to accepting how uncertain future scenarios are. Uncertainty analysis and scenario evaluation can be taken into account during land-use planning processes and evaluation of natural resources. It is common now to describe events as probable and possible, referring to various degrees of probability (Rubin 2004). Stochastic hydrology theories and approaches have developed considerably during the last 30 years. However, stochastic methods have not been used very frequently in hydrological applications, despite the impressive success of essentially the same approaches in other scientific disciplines like atmospheric modeling, meteorological forecast, ocean monitoring and ecological assessment (Christakos 2004). So far, their application to real-world problems has been limited and they have not become a routine tool for hydrological modeling (Dagan 2002; Zhang and Zhang 2004; Rubin 2004; Renard 2007). One reason is that these methods generally fail when trying to represent complexities inherent to processes that cannot be modeled without rigorous rules. There is a need to balance these rigid theories and their application with quantitative accuracy in field engineering problems involving the subsurface environment (Hunt and Doherty 2011; Ginn 2004). Renard (2007) presented a survey of publications in the field of stochastic hydrogeology over the last two decades. While the total number of publications has increased in all fields of research, the author was surprised by the observation that the proportion of stochastic hydrogeology papers remained constant,

Received: 9 May 2011 / Accepted: 8 June 2012
Published online: 28 June 2012

© Springer-Verlag 2012

R. L. Manzione (✉)
UNESP/São Paulo State University,
Campus of Ourinhos. Av. Vitalina Marcusso, 1500,
CEP: 19910-206, Ourinhos, SP, Brazil
e-mail: rlmanzione@gmail.com
Tel.: +55-14-33025716

E. Wendland
EESC/Department of Hydraulics and Sanitary Engineering,
USP/University of São Paulo, PO Box 359 CEP: 13560-970,
São Carlos, SP, Brazil
e-mail: ew@sc.usp.br

D. H. Tanikawa
UNESP/São Paulo State University,
Campus of Ourinhos. Av. Vitalina Marcusso, 1500,
CEP: 19910-206, Ourinhos, SP, Brazil
e-mail: tanikawa@hotmail.com

suggesting that interest in these techniques within the research community is not increasing. Pappenberger and Beven (2006) discuss reasons why uncertainty analysis should not be used and Clarke (2010) points out the misuse of statistical methods in reporting hydro-climatological data. Neuman (2004) believes that the level of mathematical understanding required by stochastic theories is far beyond that of most hydrogeologists, in academia and in practice. Sudicky (2004) agrees, and adds that regulations are based on single numbers, and the uncertainty in the prediction of such numbers (e.g. the probability of exceeding it) is not something regulators can readily deal with. Nourani et al. (2011) believe that problems in data interpretation given by the lack of strong predictive tools (or lack of experienced users of those tools) contribute to a failure to reach consensus about the need for key water-management actions.

The motivation of this work is to present an integration of stochastic methods based on time-series modeling and geostatistics applied to field data collection of water heads and climatological observations, in order to provide reliable information for decision making in groundwater management and land-use planning. The aim is to demonstrate, through a case study, the use of stochastic approaches to explain observed phenomena, and communicate the results in a clear and understandable form, via simulated scenarios of variations in expected water-table levels. The study area is an outcrop zone of the Guarani Aquifer System (GAS) in Brazil, one of the major water reservoirs in the world, although little is known about its dynamics and recharge processes.

Materials and methods

The nature of the problem

The GAS is a transboundary aquifer covering up to 1.2 million km² of the Argentinean, Brazilian, Paraguayan and Uruguayan territories, encompassing almost all the Paraná and Chaco-Paraná sedimentary basins. During the past decade, efforts from the four countries were concentrated on investigating this reservoir and obtaining knowledge about its recharge and discharge mechanisms, in order to achieve sustainable and optimal exploitation of the groundwater. Due to the strategic, social and economic importance of this aquifer to the four countries, it is necessary that use of this water resource is coordinated (OAS 2009).

The implementation of a monitoring system, which allows the acquisition of detailed information about the condition and behaviour at specific sites within the GAS, is a prerequisite for efficient management of the aquifer (Wendland et al. 2007). Stochastic approaches applied to reservoirs as large as the GAS can give some information to decision makers and water commissioners, which allows them to use numerical intervals in risk analysis to solve real groundwater problems. Predicting water levels at unvisited locations and predicting extreme levels on specific dates allows inferences to be made in both spatial and temporal dimensions of a physically based water system (Manziona et al. 2010). There

are still many gaps in monitoring network coverage for water levels and quality in the GAS, and there is also a lack of local knowledge about aquifer dynamics. With such general deficiency in data acquisition, stochastic methods represent an alternative way of dealing with uncertainties about water quantification. Accounting for model uncertainty is important for sustainable management of the enormous area covered by the GAS, and for the protection of its recharge areas.

Study area

The GAS was formed during the Jurassic (Botucatu Formation) and Triassic (Piramboia Formation) periods. The Piramboia Formation consists of silty-clayish sandstones of aeolian and restricted fluvial origins, and the overlying Botucatu Formation consists of well-sorted sandstones of aeolian origin (Sracek and Hirata 2002). The Triassic sandstones usually have a larger amount of clay in their lower layers, which diminishes, in relative terms, their hydraulic efficiency (Rabelo and Wendland 2009). The sandstone outcrops on the borders of the Paraná sedimentary basin are responsible for most of the aquifer recharge, and the sandstone is mostly confined by the Serra Geral basalt layer (Cretaceous). The region chosen for the present monitoring work is the Onça Creek watershed, which is located in the outcrop area of the GAS, in the central region of the province of São Paulo, Brazil (Fig. 1). This area of 5,800 ha possesses representative characteristics for the outcrop zone of the GAS and has land uses that are typical of São Paulo province, like sugarcane, reforestation (eucalyptus), citrus, pasture, and some patches of natural Cerrado vegetation, a extensive woodland-savannah present in several regions of South America. The Onça Creek watershed is located between 22°10' and 22°15' south and 47°55' and 48°00' west, being an affluent of Jacaré-Guaçú River. The Onça Creek flows mainly over sandstone of the Botucatu Formation, while at the basin outlet it flows over the Botucatu-basalt complex. Cenozoic soils present in the area result from sandstone weathering, showing homogeneous composition with almost no loam (Wendland et al. 2007). Climatic classification for the region, following Koeppen, indicates a humid subtropical climate (Cwa) with summer rains, showing a variation to a tropical climate with dry winter (Wendland et al. 2007). The mean annual precipitation is about 1,300–1,400 mm, while the annual mean temperature in the region is 20.5 °C.

Data sets from the Onça Creek watershed

The Onça Creek watershed is equipped with 23 monitoring wells (Fig. 1, with four wells outside the watershed boundaries), three compact climatological stations and a river station installed in the watershed for groundwater, weather and creek-level monitoring. Water-table depths were observed manually every 15 days, from April 2004 until July 2011, totalling more than 7 years of monitoring. These wells were deliberately selected to cover the range

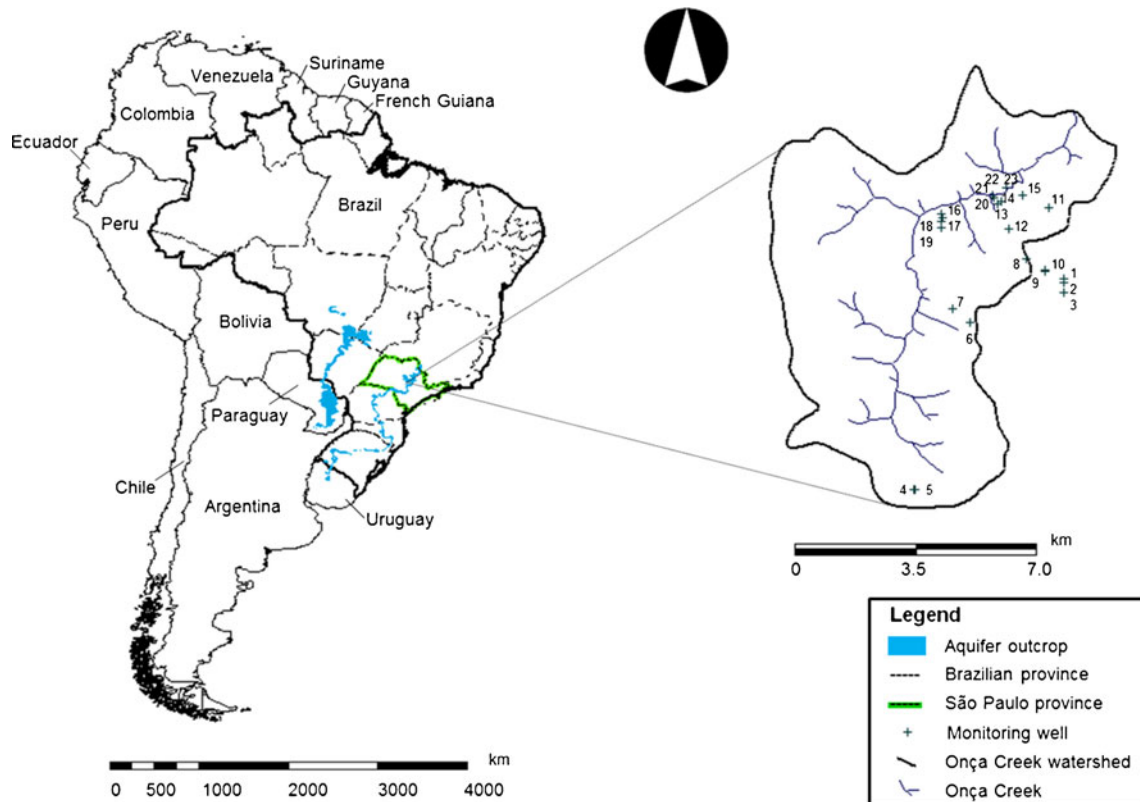


Fig. 1 Map of South America showing the Guarani Aquifer System outcrop areas and inset map of the Onça Creek watershed, located in the central part of São Paulo province, Brazil

of land uses and possible hydrogeological domains in the area, in an attempt to characterize the different responses of water-table depths in the watershed. The length of the time series of water-table depths is 2,638 days. The filter levels of the wells vary with soil depth, but are mostly around 25 m below ground level. Also, 35-year datasets for precipitation and potential evapotranspiration were available from CRHEA/USP (Centre for Water Resources and Applied Ecology of the University of São Paulo), where climatologic data are collected continuously. These data were available from 1974 until the present date, with a daily frequency. The watershed is located approximately 1.5 km away from the CRHEA station.

Combining time-series models with geostatistics on monitoring data

The characteristics of hydrogeologic systems can vary greatly in space and time, but they are usually sparsely sampled. Knowledge of system parameters is therefore partial at best, and the most that can usually be done is to quantify uncertainty through stochastic, or related, models (Winter 2004). Time-series modeling provides an empirical stochastic method to model monitoring data from observation wells, without the complexity of physical mechanistic models. Also, the stochastic component in the model allows model uncertainty to be taken into account. In time-series analysis, many authors refer to transfer function-noise models to describe the dynamic

relationship between climatological inputs and water-table depths (Box and Jenkins 1976; Hipel and McLeod 1994; Tankersley and Graham 1994; Van Geer and Zuur 1997; Yi and Lee 2003; Von Asmuth and Knotters 2004; Von Asmuth et al. 2008). In the same way, geostatistical methods are used to make probabilistic statements about quantities of interest at non-measured locations (Kitanidis 1997; Pebesma 2006). In general, these studies measure an attribute in relation to its spatial coordinates, but other GIS datasets can provide additional information for spatial prediction purposes, entering the prediction equations as predictors in a regression model, or as correlated measures. Pebesma (2006) states that in an environmental context, policy makers may be interested in aggregated, regional estimates from certain smaller or larger regions.

Stochastic simulation of time-series models and calculation of groundwater statistics

The first step in time-series analysis is to identify a model able to describe the time processes (Hipel and McLeod 1994). Ideally, the time-series model should not be based just on statistical methods, but should also include some physical insight about water-table dynamics. The PIRFICT model is an alternative to discrete-time transfer function-noise models (Von Asmuth et al. 2002). PIRFICT is an acronym for “predefined IR function in continuous time”. The predefined IR (impulse response) function is a Pearson III distribution function (PIII df), which describes the model response (water-table variation) to the system inputs. In addition, the PIRFICT

model has a stochastic component (noise term) that accounts for model uncertainty. The parameters of the adjusted IR functions have a physical meaning, since they represent the influence of drainage resistance, storage coefficient of the soil and the dispersion time of precipitation through the unsaturated zone (Von Asmuth and Knotters 2004).

In the simple case of an undisturbed phreatic system, the relationship between water-table dynamics and system inputs can be modelled as:

$$h(t) = \sum_{i=1}^N h_i^*(t) + d + r(t) \quad (1)$$

$$h_i^*(t) = \int_{-\infty}^t R_i(\tau)\theta_i(t-\tau)d\tau \quad (2)$$

$$r(t) = \int_{-\infty}^t \varphi(t-\tau)dW(\tau) \quad (3)$$

where h [L] is the observed groundwater level at time t [T], h_i^* [L] is the predicted groundwater-level change, written as a convolution integral, at time t credited to stress i , d [L] is the groundwater level corresponding to the local drainage level estimated from the data, and r represents the residuals series [L]. In the second and third equations, R_i [L/T] is the value of stress i at time τ , θ is the IR function [-], ϕ is the noise IR function [-] and W [L] is a continuous white noise (Wiener) process with properties $E\{dW(t)\}=0$, $E[\{dW(t)\}^2]=dt$, $E[dW(t_1)dW(t_2)]=0$, $t_1 \neq t_2$.

Von Asmuth et al. (2008) also introduced a more complex approach using multiple input stress series. They distinguished several types of stress, including precipitation, evapotranspiration, groundwater withdrawal (or injection), surface-water levels, barometric pressure and hydrological interventions. In the present study, $h_i^*(t)$ is credited to the precipitation and evapotranspiration stresses, modelled as transfer function noise by the PIRFICT model as follows:

$$h_p^*(t) = \int_{-\infty}^t p(\tau)\theta_p(t-\tau)d\tau \quad (4)$$

$$h_{et}^*(t) = \int_{-\infty}^t -et(\tau)f\theta_p(t-\tau)d\tau \quad (5)$$

where p is the precipitation intensity at time t [L/T], et is the evapotranspiration intensity at time t [L/T], and f is a

reduction function for et [-] depending on soil and land cover. The effect of evapotranspiration et on the groundwater levels is the same as the effect of precipitation p , but negative. When Eqs. 4 and 5 are summed as $h_i^*(t)$ in Eq. 1, evapotranspiration influence is subtracted from precipitation, similar to a water balance. The calibrated IR function θ and the noise IR function ϕ are responsible for describing the local hydrogeological conditions found at each well location. An important advantage in the use of the PIRFICT model compared with discrete-time transfer function-noise models (Von Asmuth and Bierkens 2005) is that it can deal with input and output series which have different observation frequencies and irregular time intervals, as found with typical monitoring data.

Time-series models, calibrated with observed water-table depths for a limited number of years, enable us to simulate series of extensive length using precipitation surplus/deficit as the input variable (Knotters and Van Walsum 1997). The stochastic simulation of the PIRFICT model is applied to simulate random realizations of the model to predict extreme water-table levels over time. The simulation generates long-term realizations, reconstructed from longer climatological series of precipitation and potential evapotranspiration (e.g. 30 years), or other hydrological records, as inputs. Statistics of groundwater levels can be estimated for the series generated by the PIRFICT model simulation. In this procedure, the PIRFICT model works as a filter in time, given by a convolutional process. Basically, the hypothesis is that signals of finite duration, transmitted through a series of linear systems, eventually assume a simple shape characterized by the IR function adjusted for the system (Mass 1994). This operation considers the shape of the PIII df adjusted from the parameters of each model. The interactions between the input stress signals and the system can be regarded as an operation in the time domain between the frequency expansion of each input signal and the frequency response of the groundwater system. The word signal means an observed and measurable phenomenon, which changes its magnitude in the course of time, and which has a property that is propagated. The signal changes its shape while passing through a system, so that it might be said that the system transforms the signal and the IR function of the composite system can be derived from those subsystems by convolution. The resulting series will represent the prevailing hydrological and climatic conditions, rather than specific meteorological circumstances during the monitoring period of water-table levels (Manziona et al. 2008). With this procedure, it is possible to avoid short-term disturbances from extreme climate events in the simulated time series, once the noise in the series has been filtered out, based on the aquifer system response to model climatological inputs (which is much slower than to climate variations). Then, probability density functions (PDF) are calculated for any date of interest in the agricultural calendar of the region to account for extreme levels. These levels are indicated by percentiles in the PDFs. The assessment of risk requires an understanding of

extreme events, which are far from average by definition (Winter 2004). The uncertainty is taken into account by the probabilities thresholds calculated from the PDFs and established in risk management by the decision makers.

In this study, random sampling from a normal distribution was applied, with March 18 and October 12 being chosen for the calculation of groundwater-level statistics. For these dates, the mean highest and mean lowest levels, respectively, were obtained from the monitoring series. To enable risk management of extreme water-table levels, two different levels of probability were calculated, one for each selected date. First, a 95 % probability level scenario was evaluated for March 18 as a measure for shallow water-table risk. A shallow water table can be a problem during the rainy season because it can stop machinery and make ploughing and planting operations impossible. Also, it can influence soil conditions, decreasing soil redox potential, increasing pH in acid soils, decreasing pH in alkaline soils and increasing conductivity and ion-exchange reactions. These modifications in the soil conditions might influence plant growth, by affecting the availability of nutrients and regulating uptake in the rhizosphere. Using the 0.95 percentile of the PDF, it can be said that in this area, the probability of the water levels being higher than the values of the resulting map is just 5 %, and 95 % for deeper water levels. The limit established for risk of shallow water-table levels on March 18 was 0.5 m below the ground surface. Second, a 5 % probability level was considered for October 12 as a measure for water-shortage risk. Using the 0.05 percentile, in this area, there is a 5 % probability that water levels will be deeper than the values of the resulting map, and 95 % for higher water levels. The limit established for water-shortage risk on October 12 was 25 m, exceeding the depth of all wells (dry wells characterize a scenario of water shortage in the area). This can be a problem during the beginning of the plants' development, affecting water availability and resulting in production losses. Mapping these extreme levels, it is possible to determine areas susceptible to extreme shallow and extreme deep levels in the watershed.

Geostatistical analysis: predicting water-table levels in time and space

The transfer function-noise models, once calibrated to a set of time series observed in various wells in an area, can have their predictions interpolated spatially, using ancillary information related to the physical basis of these models (Knotters and Bierkens 2001). Ancillary information can be used to improve model predictions since there is a strong relationship between the secondary variables and the target variable. Also, using geostatistical interpolators is another approach to aggregating physical knowledge about the processes into the statistical model. When abundant ancillary information related to the processes is available, it is possible to not only improve the accuracy of the spatial predictions, but also yield more plausible spatial patterns and enhance the physical

meaning of the maps, using digital elevation models (DEM) and/or classified satellite images (Hengl et al. 2007; Manzione et al. 2010; Peeters et al. 2010).

In this approach, the spatial differences in water-table dynamics are determined by the spatial variation in the system properties, while its temporal variation is driven by the dynamics of the input into the system. In this study case, the values of simulated water-table levels are interpolated spatially using geostatistical techniques (Pebesma 2004). The variability of n observations $Z(s)$, with s denoting spatial location, is the sum of a trend $m(s)$ and a residual $e(s)$:

$$Z(\mathbf{s}_i) = m(\mathbf{s}_i) + e(\mathbf{s}_i); \quad (6)$$

$$i = 1, \dots, n$$

where the trend $m(s)$ is a deterministic linear function of $s + 1$ unknown constants β_i and known covariates $f_i(s)$, fitted by linear-regression analysis:

$$m(\mathbf{s}_i) = \beta_0 + \beta_1 \cdot f_1(\mathbf{s}_i) + \dots + \beta_p \cdot f_p(\mathbf{s}_i) + e(\mathbf{s}_i) \quad (7)$$

and the residual $e(s)$ is a zero-mean random variable with stationary covariance, dependent only on separation vectors $\mathbf{s}_i - \mathbf{s}_j$ interpolated using kriging.

Complete coverage of $f(s)$ in the database for predicting $Z(s_0)$ at unobserved locations s_0 of Z , can help to use all information available optimally, find satisfactory agreement between model and data and yield adequate predictions (Pebesma 2006). The universal kriging predictor of $Z(s_0)$ is:

$$\widehat{Z}(s_0) = \sum_{k=0}^p \widehat{\beta}_k \cdot f_k(s_0) + \sum_{i=1}^n \lambda_i \cdot e(\mathbf{s}_i) \quad (8)$$

$$f_0(\mathbf{s}_i) = 1$$

The β_k are trend model coefficients estimated using Generalized Least Squares (GLS), $f_k(s_0)$ is the k -th external explanatory variable (predictor) at location s_0 , p is the number of predictors, λ_i are the kriging weights and $e(s_i)$ are the zero-mean regression residuals. The model is considered to be the best linear predictor for spatial data (Christensen 2001). The zero-mean spatially correlated residual has its spatial-correlation structure characterized by a semivariogram. The semivariogram is the spatial estimator of dependence between observation points and provides information to the kriging system when performing spatial interpolation. The results of spatial interpolation are evaluated using cross-validation (Chilès and Delfiner 1999; Pebesma 2004).

The DEM with 10-m resolution was used as ancillary information to predict water-table depths for the whole Onça Creek watershed (Fig. 2). The resulting maps can be used to evaluate possible states of the phreatic surface or risk areas for extreme deep or shallow water-table levels during any period of the year. This is valuable information for decision making and water-management policies, in order to balance economic and ecological purposes in

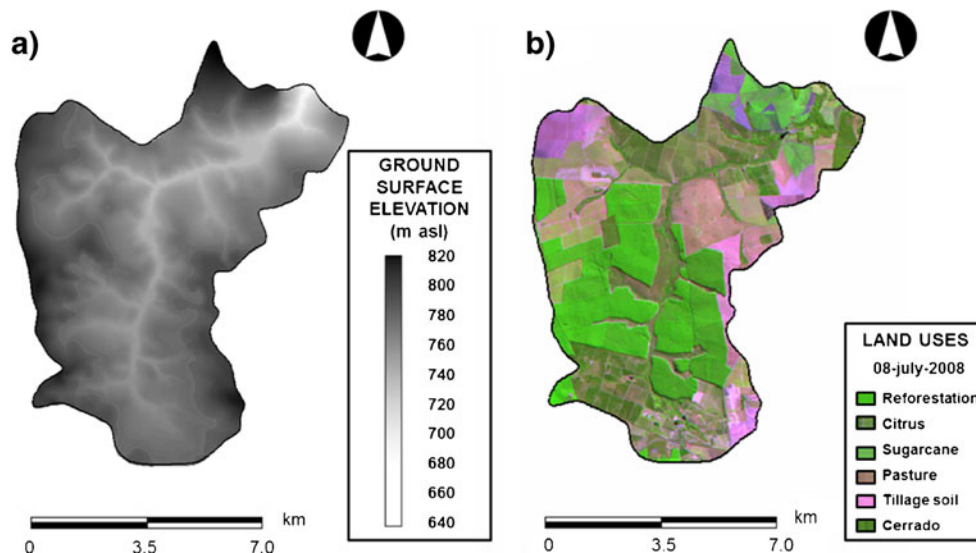


Fig. 2 a) Digital elevation model with 10-m resolution in m above sea level (asl); and b) Landsat 5 R(5)G(4)B(3) image composition for 2008's land use for the Onça Creek watershed

groundwater exploitation. The method presented has demonstrated potential for unconfined and porous aquifers in Brazil. Manzione et al. (2010) presented the results of stochastic modeling in the Brazilian Cerrados using maps with physical meaning determined by the relationship between water-table levels and elevation, considering them more understandable for the general public, rather than tables containing statistical intervals.

Results and discussion

Calibration and simulation of time-series models in the Onça Creek watershed

The calibrations and simulations of the PIRFICT model were performed for 21 wells. Two wells of the monitoring network were removed from the analysis due problems in the wells and discontinued monitoring (wells 1 and 6). The calibration statistics calculated for these 21 wells are presented in Table 1.

The calibrations suggest reasonable model performance, with an average percentage of the variance explained by the models (EVP) of 78.80 %; average root mean square error (RMSE) of 0.530 m; and average root mean squared innovation (RMSI) of 0.271 m. The RMSI values represent the mean model error due to time lag (t versus $t - 1$), and is understood as a measure of model accuracy (Von Asmuth and Bierkens 2005). Since the RMSI values are smaller than the RMSE values, the PIRFICT-model assumptions are true, because the model preserves the memory of previous time steps in order to adjust the subsequent predictions to actual observations. Considering the length of the time series, with 7 years of monitoring data, the results are satisfactory. In some wells, the results were affected by changes in land use. Well 8, for example, had the land use changed from reforestation with eucalyptus to sugar cane. The water demand from these fields is different, and in this case, not only

precipitation and evapotranspiration can influence the response of the aquifer system. Wells 9 and 10 are located under reforestation with eucalyptus. During the monitoring period the forest was cleared, and without the trees, the water levels rose until the plants sprouted again and started to intercept the meteoric water that reaches the soil and infiltrates. In these cases, Von Asmuth et al. (2008) recommend introducing other inputs into the PIRFICT model in order to characterize the aquifer response due to multiple stresses.

Figures 3 and 4 present the calibrated models for wells 16, 17, 18 and 19, and their adjusted IR functions,

Table 1 Statistics of PIRFICT-model calibrations for 21 monitoring wells in the Onça Creek watershed from April 2004 to July 2011

Well	EVP (%)	RMSE (m)	RMSI (m)
2	75.02	1.026	0.479
3	84.55	0.230	0.117
4	90.69	0.221	0.085
5	87.62	0.281	0.155
7	73.22	0.780	0.325
8	69.78	1.137	0.538
9	62.18	1.285	0.350
10	68.00	0.777	0.210
11	86.20	0.297	0.236
12	78.20	0.809	0.473
13	78.27	0.512	0.319
14	79.64	0.492	0.310
15	82.30	0.409	0.180
16	86.31	0.260	0.148
17	79.16	0.568	0.355
18	75.87	0.804	0.597
19	80.74	0.762	0.422
20	79.41	0.239	0.136
21	68.14	0.099	0.102
22	82.63	0.128	0.087
23	86.95	0.063	0.063
Mean	78.80	0.530	0.271
SD	7.43	0.357	0.162

EVP percentage of explained variance; RMSE root mean squared error; RMSI root mean squared innovation; SD standard deviation

respectively. These wells are located in a transect, where well 16 is the closest to the drainage and well 19 the more distant, with wells 17 and 18 between them. The levels are more superficial and react faster close to the river, as can be seen in the behaviour of well 16. More distant from the river, the soil coverage gets thicker and the memory of the aquifer system increases.

Once the dynamic relationship between precipitation surplus and water-table levels was established and calibrated for all wells, the PIRFICT-model was simulated more than 1,000 times. The resulting series are 1,000 realizations of the PIRFICT-model with 30-year lengths, which were sampled for the dates March 18 and October 12. The procedure reconstructed the series of water-table levels for the unobserved period, based on the available precipitation and evapotranspiration climatological series with longer length. The simulated series preserved the seasonal patterns of water-table oscillation in each of the wells, and the confidence intervals of each series were narrow, denoting a good performance for the simulation.

Mapping simulated water-table levels for selected dates

To estimate the water-table levels for the whole area of the Onça Creek watershed, the spatial correlation structure of the selected percentiles of the simulated time series for March 18 and October 12 are characterized by semivariograms with a linear trend given by elevation. The interpolation considered a linear combination of the observed values and a deterministic trend derived from the DEM with 10-m resolution in the universal kriging system. The linear correlation between elevation values and the target variable predicted water-table levels determined by the coefficient of determination (r^2) for March 18 was 0.78 and for October 12 was 0.76. This confirms the hypothesis that in the upper elevations, the

water levels are deeper than in the lower elevations (close to the drainage), where the levels are shallow, following the geomorphological patterns of the watershed (Kitanidis 1997). The relationship between variables not only improved the estimations but also aggregated information for the western part of the watershed where there were no wells (see Fig. 1). The spatial structure modelled for both scenarios was similar. For March 18, a spherical model was used with 0.5 nugget effect, 20.5 sill and 1,000-m range, and for October 12, a spherical model was used with 0.5 nugget effect, 21.5 sill and 1000-m range. The semivariance decreased 80 and 71 % for the two models, respectively, adding elevation as ancillary information to estimate water-table levels when compared with semivariograms using only water-level observations to model the spatial structure of the simulated levels.

The maps with the 0.95 percentile of the water-table level PDF simulated for March 18 and detected risk areas of shallow water-table levels are presented in Fig. 5. In wet years, the map of 0.95 percentiles can be used, for instance, to select areas with risk of shallow levels that can stop machinery and delay field operations (Manziona et al. 2010). In the case of March 18, the two risk spots shown in Fig. 5b correspond to a wetland in the outlet of the watershed where no agricultural practice takes place and to a barrage built for irrigation purposes. The maps with the 0.05 percentile of the water-table level PDF simulated for October 12 and detected risk areas of water-table levels deeper than 25 m are presented in Fig. 6. In dry years, the map of 0.05 percentiles can be used to select areas with risk of water shortage and dry wells (Manziona et al. 2008). In Fig. 6, it is possible to detect four spots with risk of levels deeper than 25 m. The spots on the borders of the watershed deserve attention because of the high uncertainty about the predictions, and the spot in the eastern part of the

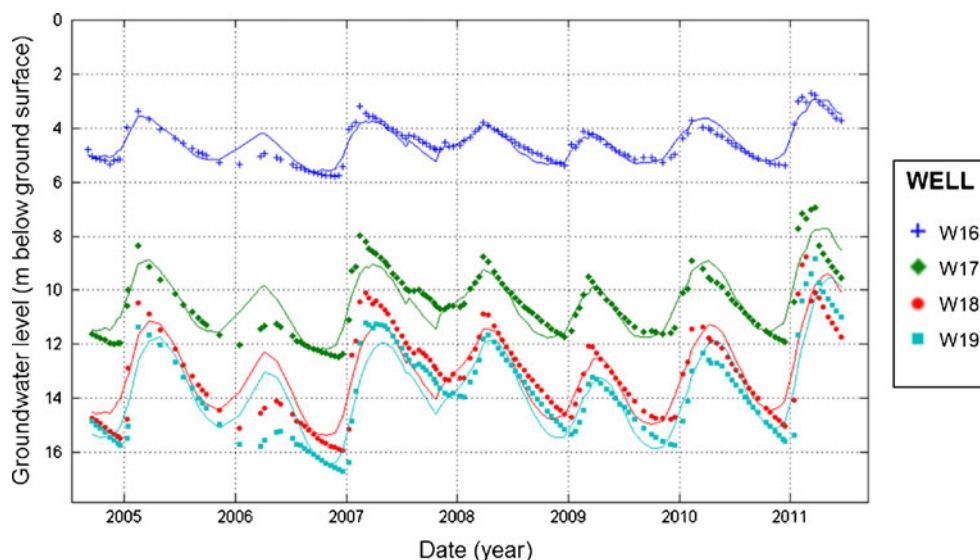


Fig. 3 Examples of PIRFICT-model calibrations for water-table level time series at wells 16, 17, 18 and 19 in the Onça Creek watershed (symbols=observations; lines=modeling results)

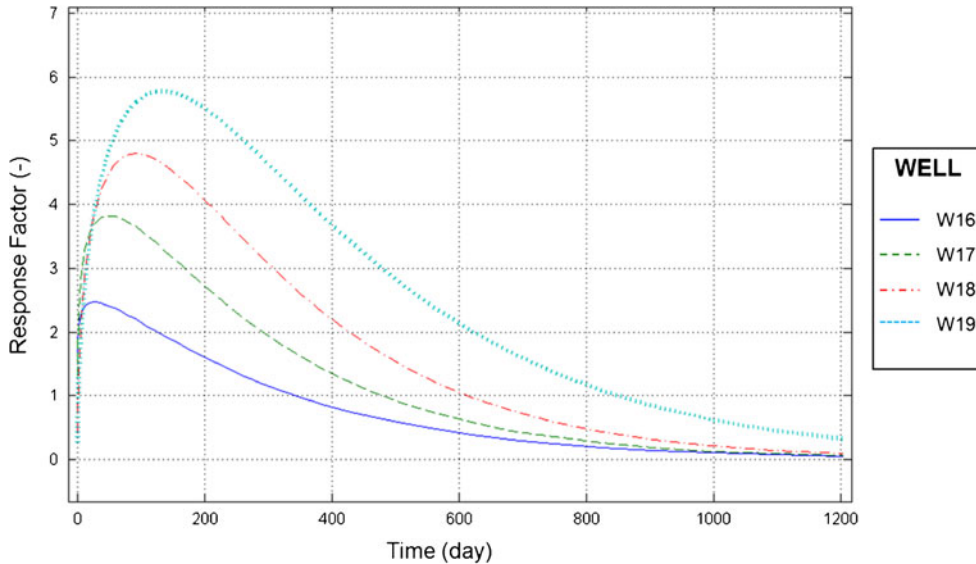


Fig. 4 Impulse response functions adjusted for the input series of precipitation, for wells 16, 17, 18 and 19 in the Onça Creek watershed

watershed because it is close to irrigated citrus fields. These limits are just examples and can be changed and adjusted for different purposes.

The scenarios presented for March 18 and October 12 contain valuable information about groundwater dynamics in the watershed, even with all the sources of uncertainty already mentioned. From these maps, water boards and land-use planners can start a discussion about strategies for the next cultivation season and/or harvest planning, based on the prevailing climatological condition and soil-water balance. It is important to mention that areas with shallow levels represent more liability than areas with deep levels, since the range of the variation is much smaller and less uncertain. It is also an indicator that efforts to improve the monitoring network should be concentrated on those areas with large uncertainty, if it is desired to improve model results. Some would say that scenarios for extreme percentiles are unrealistic, since it

would be impossible to imagine that in all spots in an area the water-table levels are reaching such extreme thresholds. However, for planning activities, these maps could reveal potential risk areas that deserve attention during land-use regulation and when assessing the resilience of agricultural activities. Resilient systems can have a higher probability of success in these uncertain times of climate change.

The results of the spatial interpolation for March 18 and October 12 were evaluated by cross-validation (Table 2). For March 18, the standard deviation (SD) values for the observations are higher than those of the predictions, pointing to smoothed interpolation values (for example, 8.38 versus 7.33). The mean predicted values reflect the mean observed water-table levels. Also, the mean interpolation errors are small (-0.04 m). The mean and standard deviation of the Z-score had values close to zero and one, respectively, characteristics of the kriging

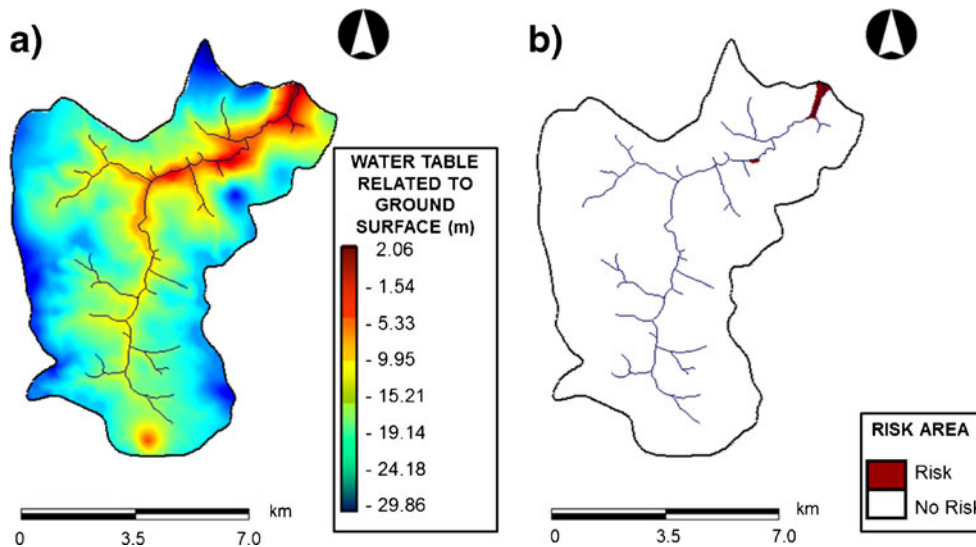


Fig. 5 a Estimated water-table depths that will be exceeded with 5 % probability; and b areas with risk of water-table depths shallower than 0.5 m; on March 18, for the Onça Creek watershed

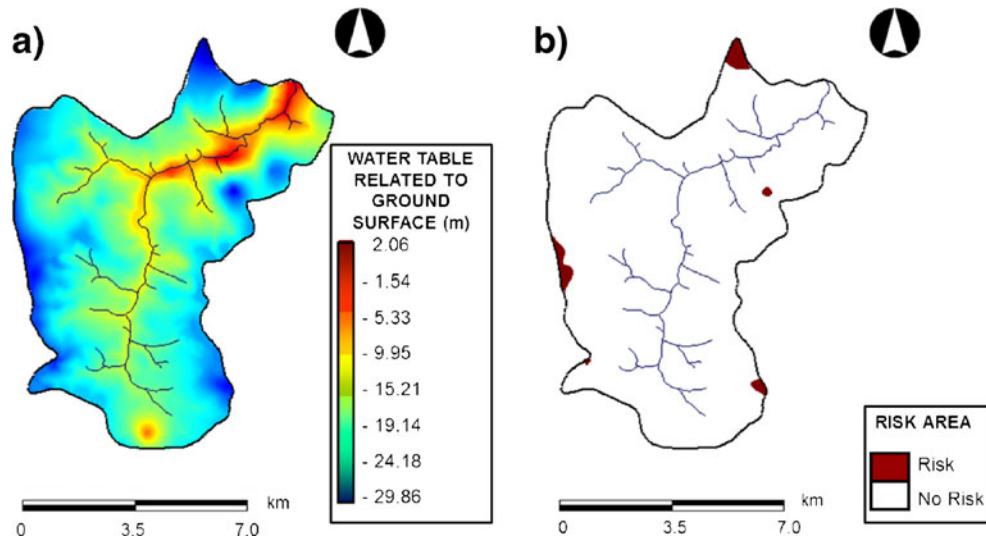


Fig. 6 a Estimated water-table depths that will be exceeded with 95 % probability; and b areas with risk of water-table depths deeper than 25 m; on October 12, for the Onça Creek watershed

system (Chilès and Delfiner 1999; Christensen 2001) as a BLUE (Best Linear Unbiased Estimator). For October 12, the SD values for the observations are also higher than those of the predictions (for example, 8.10 versus 6.89). The mean predicted values reflect the mean observed water-table levels and the mean interpolation errors are small (-0.01 m). The mean and standard deviation of the Z-score had values close to zero and one, respectively. The results for both dates suggest good performances of the kriging system and the estimations.

Due to the lack of information, to obtain better results in the spatial interpolation, the degree of uncertainty in the maps should be considered for further analysis. The sources of uncertainty can be explained as resulting from different sources: uncertainties related to the data (observed water-table levels, climatic database, measured and estimated quantities); uncertainty associated with time-series modeling (model calibrations); and with the model of spatial variation (lack of data to cover all the watershed and description of the spatial structure of the phenomena). However, understanding these errors gives a margin for water planning; for example, installing longer cables and connections during well construction and positioning the water pump 10 m deeper for operation. Also, for more than half of the watershed area, there was no

information about the water-table dynamics. Now it is possible to imagine its behavior from the geomorphology of the region, based on the physical meaning of the final maps.

Integration of stochastic methods as a framework to predict water-table levels for GAS

This kind of integration of methods deserves attention because it can offer a framework for establishing a dialogue between basic research and applied technology (Christakos 2004). Stochastic models can capture the evolution of risk as a dynamic concept, essentially a probabilistic concept that changes in space and time (Winter 2004). For Freeze (2004), there is a need to adjust the complexity of modeling with decision-makers expectations, and, he remarks that uncertainty analysis feeds risk analysis and consequently economic analysis. In developing counties like Brazil, where data scarcity is often a problem, sometimes model simplicity is required to see modelling results applied in decision systems as routine. It helps clients to select the most cost-efficient strategies for their purposes. Pappenberger and Beven (2006) said that users need predictions for future decisions based on an assessment of potential future states and the risk associated with them, products of uncertainty estimation. For Neuman (2004), real-world situations must be analyzed using methods that incorporate realistic combinations of randomness (stochasticity) and determinism (via conditioning). Dagan (2002), Neuman (2004) and Renard (2007) highlight the promotion of stochastic software to facilitate application of these concepts by practitioners. In addition, it is recommended that visualizations of evaluated scenarios are provided, to enhance the interpretation of the results via maps. The fact that many decision makers cannot deal with uncertainty, a topic addressed by Pappenberger and Beven (2006), Christakos (2004), Freeze (2004), Sudicky (2004), Renard

Table 2 Summary of cross-validation for spatial interpolation of water-table levels at March 18 and October 12

	Observed	Predicted	Pred. – Obs.	Pred. SD	Z-score
March 18					
Mean	-10.44	-10.44	-0.04	3.07	-0.02
SD	8.38	7.33	5.26	1.37	1.27
October 12					
Mean	-11.49	-11.49	-0.01	3.13	-0.01
SD	8.10	6.89	5.36	1.41	1.30

Pred. – Obs. predicted minus observed values; SD standard deviation; Z-score = (Pred – Obs)/kriging variance. All units: meters

(2007) and others, motivates efforts to familiarize the general public with probabilistic concepts and to develop tools which allow the results to be visualized. In this process, is necessary to continue and increase the communication of these kinds of results from academia and from professionals familiar with stochastic theories (Ginn 2004; Renard 2007). Maps containing physical information about the underlying problem are important, since the visualization of groundwater scenarios is difficult for many people. The degree of uncertainty contained in a map should be evaluated and taken into account. Depending on the purpose of the study, the level of uncertainty can be established and used as a risk measure.

The Brazilian government and local authorities are making several efforts to improve the GAS monitoring network coverage for hydro-climatological purposes. In the near future, a wide range of time series will be available for practitioners, so that they can analyse the data for specific studies. The results of the environmental protection and sustainable development of the Guarani Aquifer System project (OAS 2009) are already available, and present new discoveries about the reservoir. New agreements have been reached to continue these studies. All sources of information can be aggregated in the decision-making processes, to help groundwater management at different scales, from the watershed scale (which forms the basis of the Brazilian National Water law), to the regional and continental domains of the GAS. In particular, for groundwater and for GAS, it is an opportunity to reset what Hunt and Doherty (2011) called the balance of theory and application. The authors ask for a renewed emphasis on applications designed to test theoretical constructs from the collaboration between theoretical and applied researchers, with training elements explicit in the work plan, not left to chance. Sustainable management and protection of GAS resources can be achieved by analysing the aquifer globally but acting locally on water-management issues. Accurate predictions will be difficult to reach for the whole aquifer domain. Local actions should be based on local analysis of aquifer dynamics and possible states or scenarios for groundwater management evaluated, discussed and considered. Stochastic methods appear to be a useful alternative, to introduce research and field monitoring results into the discussion meetings. Understanding model uncertainty and experimenting with these results is desirable for this purpose.

Conclusions

Time-series analysis, combined with geostatistics, modelled the dynamic behaviour of water-table levels in the Onça Creek watershed, resulting in maps with possible water-table levels for specific dates. The PIRFICT model described different hydrological behaviours within a watershed, from the same data set, and simulated water levels for the selected dates of March 18 and October 12. Stochastic hydrogeology applications can provide useful information for real-world application and assist the decision-making processes in groundwater management and long-term water policy. The visualization of the final

results in the form of maps enhanced the comprehension of the stochastic results, as is routine in land-use planning and evaluation of natural resources. For large aquifers like the GAS, it is important to explore and aggregate different sources of information when modeling monitoring data, and to consider uncertainty in the decision-making processes for groundwater management and quantitative estimations.

Acknowledgments The authors are grateful to São Paulo Research Foundation (FAPESP) for financial support (Projects n. 2009/05204-8, 2010/14914-6, 2011/11484-3) and to The Brazilian National Council for Scientific and Technological Development (CNPq) for infrastructural support.

References

- Box GEP, Jenkins GM (1976) Time series analysis: forecasting and control. Holden-Day, San Francisco
- Chilès JP, Delfiner P (1999) Geostatistics: modeling spatial uncertainty. Wiley, New York
- Christakos G (2004) A sociological approach to the state of stochastic hydrogeology. *Stoch Environ Res Risk Assess* 18:274–277. doi:10.1007/s00477-004-0197-1
- Christensen R (2001) Linear models for multivariate, time series, and spatial data, 2nd edn. Springer, New York
- Clarke RT (2010) On the (mis)use of statistical methods in hydro-climatological research. *Hydrol Sci J* 55:139–144. doi:10.1080/02626661003616819
- Dagan G (2002) An overview of stochastic modeling of groundwater flow and transport: from theory to applications. *EOS Trans AGU* 83:621. doi:10.1029/2002EO000421
- Freeze RA (2004) The role of stochastic hydrogeological modeling in real-world engineering applications. *Stoch Environ Res Risk Assess* 18:286–289. doi:10.1007/s00477-004-0194-4
- Ginn TR (2004) On the application of stochastic approaches in hydrogeology. *Stoch Environ Res Risk Assess* 18:282–284. doi:10.1007/s00477-004-0199-z
- Hengl T, Heuvelink GBM, Rossiter DG (2007) About regression-kriging: from equations to case studies. *Comput Geosci* 33(10):1301–1315. doi:10.1016/j.cageo.2007.05.001
- Hipel KW, McLeod AI (1994) Time series modeling of water resources and environmental systems. Elsevier, Amsterdam
- Hunt RJ, Doherty J (2011) Interesting or important? Resetting the balance of theory and application. *Ground Water* 49(3):301. doi:10.1111/j.1745-6584.2011.00807.x
- Kitanidis PK (1997) Introduction to geostatistics: applications in hydrogeology. Cambridge University Press, Cambridge, UK
- Knotters M, Bierkens MFP (2001) Predicting water table depths in space and time using a regionalised time series model. *Geoderma* 103:51–77. doi:10.1016/S0016-7061(01)00069-6
- Knotters M, Van Walsum PEV (1997) Estimating fluctuation quantities from time series of water-table depths using models with a stochastic component. *J Hydrol* 197:25–46. doi:10.1016/S0022-1694(96)03278-7
- Manziane RL, Knotters M, Heuvelink GMB, Von Asmuth JR, Câmara G (2008) Predictive risk mapping of water table depths in a Brazilian Cerrado area. In: Stein A, Wenzhong S, Bijker W (eds) Quality aspects in spatial data mining. CRC, Boca Raton, FL doi:10.1201/9781420069273.ch7
- Manziane RL, Knotters M, Heuvelink GMB, Von Asmuth JR, Câmara G (2010) Transfer function-noise modeling and spatial interpolation to evaluate the risk of extreme (shallow) water-table levels in the Brazilian Cerrados. *Hydrogeol J* 18(8):1927–1937. doi:10.1007/s10040-010-0654-5
- Mass C (1994) On convolutional processes and dispersive groundwater flow. PhD Thesis, Delft University of Technology, The Netherlands

- Neuman SP (2004) Stochastic groundwater models in practice. *Stoch Environ Res Risk Assess* 18:268–270. doi:10.1007/s00477-004-0192-6
- Nourani V, Ejlali RG, Alami MT (2011) Spatiotemporal groundwater level forecasting in coastal aquifers by hybrid neural networks-geostatistics model: a case study. *Environ Eng Sci* 28(3):217–228. doi:10.1089/ees.2010.0174
- OAS (Organization of American States) (2009) Guarani aquifer: strategic action plan. Available via: http://www.ana.gov.br/bibliotecavirtual/arquivos/20100223172013_PEA_GUARA-NI_Ing.pdf Cited 20 April 2011
- Pappenberger F, Beven KJ (2006) Ignorance is bliss: or seven reasons not to use uncertainty analysis. *Water Resour Res* 42:W05302. doi:10.1029/2005WR004820
- Pebesma EJ (2004) Multivariable geostatistics in S: the Gstat package. *Comput Geosci* 30:683–691. doi:10.1016/j.cageo.2004.03.012
- Pebesma EJ (2006) The role of external variables and GIS databases in geostatistical analysis. *Trans GIS* 10(4):615–632. doi:10.1111/j.1467-9671.2006.01015.x
- Peeters L, Fasbender D, Batelaan O, Dassargues A (2010) Bayesian data fusion for water table interpolation: incorporating a hydrogeological conceptual model in kriging. *Water Resour Res* 46:W08532. doi:10.1029/2009WR008353
- Rabelo J, Wendland E (2009) Assessment of groundwater recharge and water fluxes of the Guarani Aquifer System, Brazil. *Hydrogeol J* 17:1733–1748. doi:10.1007/s10040-009-0462-y
- Renard P (2007) Stochastic hydrogeology: What professionals really need? *Ground Water* 45:531–541. doi:10.1111/j.1754-6584.2007.00340.x
- Rubin Y (2004) Stochastic hydrogeology: challenges and misconceptions. *Stoch Environ Res Risk Assess* 18:280–81. doi:10.1007/s00477-004-0193-5
- Srceck O, Hirata R (2002) Geochemical and stable isotopic evolution of the Guarani Aquifer System in the state of São Paulo, Brazil. *Hydrogeol J* 10:643–655. doi:10.1007/s10040-002-0222-8
- Sudicky E (2004) On certain stochastic hydrology issues. *Stoch Environ Res Risk Assess* 18:285. doi:10.1007/s00477-004-0196-2
- Tankersley CD, Graham WD (1994) Development of an optimal control system for maintaining minimum groundwater levels. *Water Resour Res* 30:3171–3181. doi:10.1029/94WR01790
- Van Geer FC, Zuur AF (1997) An extension of Box-Jenkins transfer/noise models for spatial interpolation of groundwater head series. *J Hydrol* 192:65–80. doi:10.1016/S0022-1694(96)03113-7
- Von Asmuth JR, Knotters M (2004) Characterising groundwater dynamics based on a system identification approach. *J Hydrol* 296:118–134. doi:10.1016/j.jhydrol.2004.03.015
- Von Asmuth JR, Bierkens MFP (2005) Modeling irregularly spaced residual series as a continuous stochastic process. *Water Resour Res* 41:W12404. doi:10.1029/2004WR003726
- Von Asmuth JR, Bierkens MFP, Maas C (2002) Transfer function noise modeling in continuous time using predefined impulse response functions. *Water Resour Res* 38:1–23. doi:10.1029/2001WR001136
- Von Asmuth JR, Maas K, Bakker M, Petersen J (2008) Modeling time series of ground water head fluctuation subjected to multiple stresses. *Ground Water* 46:30–40. doi:10.1111/j.1745-6584.2007.00382.x
- Wendland E, Barreto CEAG, Gomes LH (2007) Water balance in the Guarani Aquifer outcrop zone based on hydrogeologic monitoring. *J Hydrol* 342:261–269. doi:10.1016/j.jhydrol.2007.05.033
- Winter CL (2004) Stochastic hydrogeology: practical alternatives exist. *Stoch Environ Res Risk Assess* 18:271–273. doi:10.1007/s00477-004-0198-0
- Yi M, Lee K (2003) Transfer function-noise modeling of irregularly observed groundwater heads using precipitation data. *J Hydrol* 288:272–287. doi:10.1016/j.jhydrol.2003.10.020
- Zhang YK, Zhang D (2004) Forum: the state of stochastic hydrology. *Stoch Environ Res Risk Assess* 18:265. doi:10.1007/s00477-004-0190-8