

Automated Time Series Modeling for Piezometers in the National Database of the Netherlands

by Willem J. Zaadnoordijk^{1,2}, Stefanie A.R. Bus³, Aris Lourens^{3,4}, and Wilbert L. Berendrecht⁵

Abstract

The Geological Survey of the Netherlands (TNO-GSN) maintains a public national database of groundwater head observations. Transfer function-noise modeling has been applied to the time series in order to extract the impulse response functions for precipitation and evaporation for each piezometer. An automated procedure has been developed to assess the quality of the time series and of the models. The time series models of sufficient quality offer far more homogeneous data on the piezometric head than the original measurements. This allows for improved mapping of the head at a specific date or of characteristics of the head like average summer or winter levels. Also, the separation of precipitation and evaporation from other influences is useful for groundwater management and policy. The individual time series models are available online with interactive graphics (<https://www.grondwatertools.nl/grondwatertools-viewer>). The spatial patterns of the impulse response function characteristics can support analyses of the groundwater system.

Introduction

The Geological Survey of the Netherlands (TNO-GSN) provides data concerning the subsurface through the website <http://www.dinoloket.nl>. Among these data are currently over 480 million observed groundwater heads from 80,000 piezometers. These time series implicitly contain information about the groundwater system, such as the response of the groundwater head to precipitation and evaporation.

The piezometric time series are inhomogeneous. The monitoring periods and frequencies vary, there are gaps, and the quality control depends on the owners

of the piezometers. More homogeneous data on the piezometric heads in the Netherlands was needed. Time series modeling has been chosen as a tool to do create these. The inclusion of precipitation and evaporation as explanatory variables in the time series modeling is beneficial in this process because these generally are the main influences on the piezometric head in the Netherlands.

The availability of more homogeneous data on the piezometric head aids analyses both in the data selection process and in the quality of the results. The former because many checks have been performed already and piezometers have been classified. The latter because the quality of the data is better and also better known.

The separation of the influence of precipitation and evaporation from other influences often also is a separation of natural and antropogenic influences. This is important for water management. Moreover, the impulse response function for precipitation is a signature of the groundwater system. Because of this, it is expected that the database with transfer function-noise models can be used to complement information from other sources like bore logs and geophysical exploration to better delineate aquifers and aquitards and improve their hydraulic parameterization.

For such applications, it is needed to automate the time series modeling and evaluation of the quality of the model. Especially, the latter poses scientific challenges in which criteria to use and how to score them.

Analysis of nationwide piezometer data is not new: for example, Russo et al. (2014) analyzed time series of

¹Water Resources Section, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, Netherlands.

²Corresponding author: TNO—Geological Survey of the Netherlands, P.O. Box 80015, 3508 TA, Utrecht, Netherlands; willem_jan.zaadnoordijk@tno.nl

³TNO—Geological Survey of the Netherlands, P.O. Box 80015, 3508 TA, Utrecht, Netherlands.

⁴Department of Physical Geography, Faculty of Geosciences, Utrecht University, Netherlands.

⁵Berendrecht Consultancy, Harderwijk, Netherlands.

Article impact statement: Time series models for piezometers in the Dutch national database are useful for individual piezometers and spatial patterns.

Received March 2018, accepted August 2018.

© 2018 The Authors. *Groundwater* published by Wiley Periodicals, Inc. on behalf of National Ground Water Association.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

doi: 10.1111/gwat.12819

24,532 piezometers across the contiguous United States. They performed trend analysis on yearly averages of piezometric heads and determined correlations between heads on the one hand and extractions and climate on the other hand. However, such analyses do not relate the changes to the quantities causing them and can give only long time trends.

During the past 30 years, groundwater head series have been frequently investigated using time series models such as transfer function-noise models (e.g., Gehrels et al. 1994; Kim et al. 2005; Manzione et al. 2010; Obergfell et al. 2013). These models include information about the groundwater response to driving forces, mainly precipitation and evaporation. Additionally, these models include a noise model for the unexplained fluctuations, the so called residuals.

The database of TNO-GSN contains a large amount of groundwater head time series. Applying transfer function-noise modeling to so many series is a challenging task. The validity of the models has to be assessed, which is usually done manually (e.g. Thyer et al. 2009; Schoups and Vrugt 2010). This leads to the question how one can model a large amount of time series and efficiently determine the quality of these models?

This paper describes the data and the applied time series modeling and evaluation method. The presentation of individual models through an interactive web interface is introduced. Applications of the information from the transfer function-noise models are presented and discussed.

Data and Methods

Data

We use the time series of piezometric heads in the DINO database. This is the national database of piezometric heads in the Netherlands, which can be accessed through the website <http://www.dinoloket.nl>. Practically all piezometers are located in sandy sediments; only some are located in the few sandstone or chalk aquifers present in the Netherlands. A commonly applied frequency for manual observations is twice per month. An increasing number of piezometers is equipped with an automatic pressure logger, usually set to a frequency of one observation per day.

The Royal Netherlands Meteorological Institute KNMI operates a meteorological monitoring network covering the entire country. Daily data is made available for 323 precipitation stations and 35 weather stations through a website and via webservices (see <http://www.knmi.nl>). At the latter, the Makkink evaporation is determined in addition to precipitation. Makkink evaporation reflects evapotranspiration from grass, which is not limited in water uptake (De Bruin and Lablans 1998). We use these precipitation and Makkink evaporation data for input in our time series modeling.

Transfer Function–Noise Models

The basis of the approach is modeling of head time series using transfer function-noise modeling with precipitation and evaporation as independent variables (Figure 1). We use a setup that has proven itself in many practical applications (see e.g., Bakker et al. 2007; Manzione et al. 2010; Peterson and Western 2014; Shapoori et al. 2015), consisting of:

- An impulse response function for precipitation which is used for convolution with the precipitation to give the transfer of the precipitation to its contribution to the piezometric head;
- An impulse response function for evaporation which is either a separately estimated function, or a factor times the function used for precipitation;
- A noise model with exponential decay.

Berendrecht and van Geer (2016) added a Kalman filter to this setup and the option to perform dynamic factor analysis on the residuals of multiple time series models.

We used this method to simulate the head with a time step of 1 day using the available precipitation and evaporation data which also has a 1 day time step.

Following Besbes and de Marsily (1984), the gamma distribution function multiplied by a factor A has been chosen as response function as displayed in Equation 1.

$$\theta_t = A \frac{a^n t^{n-1} e^{-at}}{\Gamma(n)} \quad (1)$$

Herein, n and a are the shape and decay rate parameter, respectively, of the gamma distribution, $\Gamma(n)$ is the gamma function, and t the time after the impulse. Parameter A is equal to the area under the impulse response function and this corresponds to the unit step response of the head to the input (precipitation or evaporation).

For numerical reasons Equation 1 is rewritten as:

$$\theta_t = A^* t^{n-1} e^{-at} \quad (2)$$

where $A^* = Aa^n/\Gamma(n)$.

The following equation describes the time series models:

$$h_t = d_t + n_t + b \quad (3)$$

where b is the base elevation of the time series model, n_t is the noise part and d_t is the deterministic part. The latter are given by:

$$d_t = \sum_{\tau=-\infty}^t (N_\tau - f_c E_\tau) A^* \tau^{n-1} e^{-a\tau} \quad (4)$$

$$n_t = \phi n_{t-1} + \eta_t \quad (5)$$

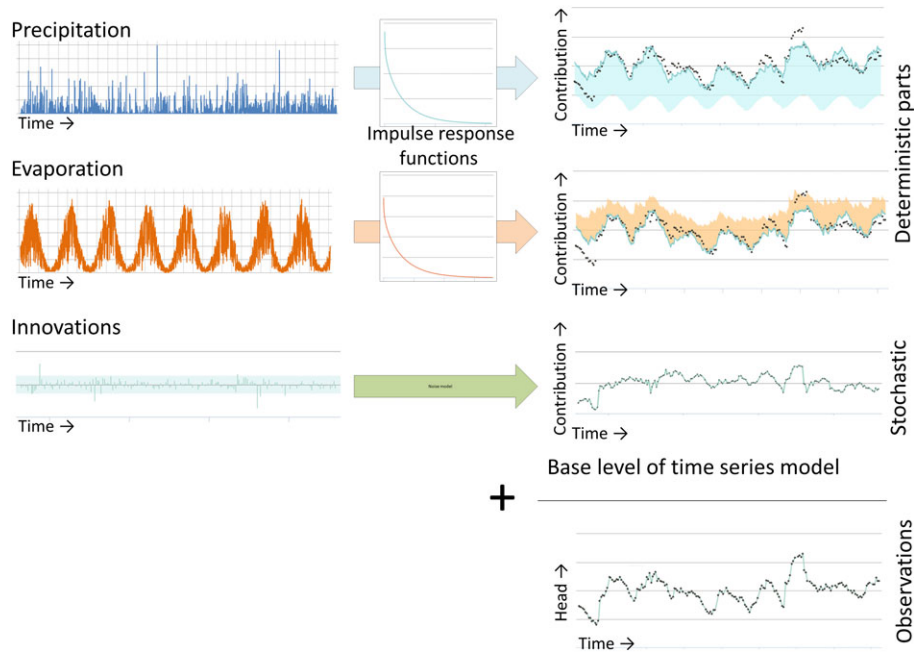


Figure 1. Setup of transfer function-noise model used for modeling head time series.

The variables N_t and E_t are the precipitation and evaporation at time t , respectively, and the parameter f_c is the ratio between the impulse response function for evaporation and precipitation. The stochastic noise process n_t is described by a first-order autoregressive model with parameter ϕ and zero mean white noise process η_t . Following Von Asmuth et al. (2002), the base elevation b has been eliminated from Equation 3 by centering the terms around their estimated means, giving:

$$h_t - \bar{h} = (d_t - \bar{d}) + (n_t - \bar{n}) \quad (6)$$

with

$$b = \bar{h} - \bar{d} - \bar{n} \quad (7)$$

The other model parameters in Equation 3, A^* , n , a , f_c , and ϕ are unknown and estimated by maximum likelihood combined with the Kalman filter for handling irregularly observed or sparse data (Berendrecht and van Geer 2016).

Automated Quality Assessment

A crucial aspect of automated calibration of time series is the quality assessment procedure. In general, there are several reasons for model inadequacy:

- Observations do not contain enough information about the groundwater dynamics (small number or short time period);
- Time series has large number of measuring errors (outliers, steps, drift);
- Strong effect from other influences than precipitation and evaporation;
- Nonlinear behavior of the groundwater system;

- Inadequacy of the model (gamma distribution function) to correctly describe the impulse response.

With a large number of series it is practically impossible to evaluate each series manually. Therefore, criteria have to be defined to judge several stages of the modeling in order to filter out time series with one or more of the issues mentioned above. These criteria need to be robust as they determine which results of the modeling will be presented on the website.

We have chosen to present model results at three levels:

- 1 Groundwater time series and associated statistics: this is simply a representation of the data with some general statistics (number of data, mean, percentiles, etc);
- 2 Components of time series that can be explained by meteorological driving forces (precipitation and evaporation);
- 3 Regime curve based on a long-term (at least 20 years) simulation with the time series model.

Level 1 (groundwater time series and associated statistics) are always presented on the website since the time series data is more or less “raw” data without any interpretation. No specific criteria are therefore used for this level.

Level 2 (components explained by precipitation and evaporation) presents a result of the time series modeling and therefore requires automated evaluation of the model results. For this level, we have defined two sets of criteria. The first set evaluates the dataset in terms of length and number of observations, in order to filter out all time series that are likely to result in inadequate time series models. These criteria are:

- At least 8 years of observations. At least several years of data are required to adequately model groundwater time series. As a robust and general rule of thumb for the Dutch situation, we applied a criterion of 8 years;
- At least 84 observations available in the last 8 years of the times series. Again this is a general rule of thumb in order to prevent modeling time series with very few observations.

In order to present information applicable to the current situation, we have chosen to include the following criterion:

- Last observation after the year 1994.

Based on these criteria we have defined the label TSOK. If all criteria are satisfied, TSOK is set to 1, otherwise it is set to 0. If TSOK = 0 no model will be calibrated for that time series (level 2 and 3 will not be evaluated).

The second set of criteria for level 2 evaluates the time series modeling result in terms of model output. We do not use the very strict criteria applied in common time series analysis, because the model is not used for any kind of prediction or whatsoever. The model is purely meant as a first indication for what part of the observed dynamics can be explained by precipitation and evaporation (as a form of regression analysis). In this context, we have defined the following criteria:

- Explained Variance R^2 greater than 0.1;
- Absolute correlation between deterministic component and residuals less than 0.3. A large correlation means that the deterministic component cannot be distinguished from the residuals. The value of 0.3 based on initial research on the dataset and may be adjusted in the future based on more detailed analysis of the results;
- Decay rate parameter a greater than 0.002. If trends or low-frequency patterns are present in a time series, the transfer function-noise model has the tendency to overfit this trend by decreasing the parameter a . Therefore, models with small values of a have a high possibility of being overfitted and are hence discarded. Similar to the correlation criterion, the value of 0.002 is more or less arbitrary and may be adjusted in the future.

If all these criteria are met, the time series is labeled with MODOK = 1; otherwise MODOK = 0 and the model will not be presented on the website (and level 3 is not evaluated).

Presentation of level 3 (regime curve) requires that the calibrated time series model may be applied for simulation. For this, some more strict criteria have been defined evaluating the predictive performance of the model:

- Explained Variance R^2 greater than 0.3. Although a value of 0.3 is still quite low, indicating that only a small part of the time series can be explained by

Table 1
Criteria for Rejecting Models

Description	MODOK	REGIMEOK
Rate parameter of distribution functions	$a < 0.002$	
Explained fraction	$R^2 < 0.1$	$R^2 < 0.3$
Correlation between stochastic and explained component	$ \rho > 0.3$	$ \rho > 0.2$
p value for autocorrelation when $R^2 < 0.8$		p value < 0.01

Table 2
Criteria for Issuing a Warning

Description	Model	Regime Curve
Gain uncertain	$A^* < 1.96 \cdot \sigma_{A^*}$	
Shape parameter uncertain	$n < 1.96 \cdot \sigma_n$	
Rate parameter uncertain	$a < 1.96 \cdot \sigma_a$	
p value when $R^2 < 0.8$		p-value < 0.05
p value when $R^2 \geq 0.8$		p-value < 0.01

the driving forces, the model can still be useful for simulation as long as the residual model satisfies the criteria below;

- Absolute correlation between deterministic component and residuals less than 0.2. Simulation of the time series is based on the assumption that there is no correlation. We have made a pragmatic choice by taking the value of 0.2 as criterion;
- Simulation is also based on the assumption that the innovations (driving force for the residual model) are serially uncorrelated. We applied the the Ljung-Box test statistic (Stoffer and Toloï 1992) to test the null hypothesis of non-correlated innovations: if the p value < 0.01 we reject the null hypothesis.

Table 1 summarizes the criteria for rejection. If at least one criterion is met then the corresponding label (MODOK, REGIMEOK) is set to 0 (indicating not OK).

As some of the criteria above are not very strict, the accepted models are further judged to determine whether we have to issue a warning at the website. A warning for the model quality is issued when at least one of the gain (A^*), the shape (n), or the rate parameter (a) has a value that is small relative to its standard deviation from the estimation. For the quality of the regime curve the p-value is tested against a threshold value. In Table 2 the warning criteria are summarized.

Depending on the label values and the number of warnings, four different quality classes are recognized (with NWARN the number of warnings):

- Insufficient (MODOK = 0);
- Potentially useful (MODOK = 1; REGIMEOK = 0);

- Decent (MODOK = 1; REGIMEOK = 1; NWARN > 0);
- Good (MODOK = 1; REGIMEOK = 1; NWARN = 0).

To limit the CPU demand for the web server, the choice has been made to store all information necessary to visualize the time series models and model statistics in a database, and only perform a calculation when the user interactively selects the option to determine a time series model for a different period. The database is updated each night when new observations have become available in the groundwater head database.

Results

Web Viewer

The models are made available via a website (<https://www.grondwatertools.nl/grondwatertools-viewer>). On an interactive map the classification of the model quality for each location is visible. For multi piezometer wells the choice is made to visualize the class of the upper piezometer. On the map, the user can select a location to view the time series and the model(s). In Figure 2 a part of such a map is shown. Depending on the classification of the quality of a time series model, the visualization is limited to the observed time series only (45% TSOK = 0 and 16% insufficient model), the observed time series together with the model (potentially useful 15%), or the observed time series, model, and regime curve (13% decent and 14% good time series models).

The website shows the location of the well on a map together with an overview of the piezometers in the well and the option to switch to another piezometer at the location.

Model Visualization

The visualization of accepted models consists of various graphs and statistics. The deterministic part of the model is shown as a daily time series in a graph together with the observations and optionally the contribution of the precipitation and evaporation to the observed dynamics. The response functions are shown graphically together with some characteristics (the unit step response M_0 , times of the average, median, and 90% response, and the time of the peak in the response). Also, the parameter values of the response function and associated standard deviations are available together with graphs of the statistical component of the time series model, the input of the statistical component (innovations), and the autocorrelations of the innovations. This provides information for the user to judge the model.

Additional information is shown for decent and good models, consisting of the regime curve and climate representative statistics. The regime curve shows the average yearly fluctuation of the head and quartile ranges around it (Figure 2). The climate representative statistics characterize the average yearly fluctuation in numbers of an average high, average low, and average spring

groundwater level—a classification which is commonly used in groundwater management in the Netherlands (Van Heesen 1970; Finke et al. 2004).

The web interface also allows the user to create new time series models of individual piezometers for different periods. This allows the user to estimate changes of the groundwater system in time.

Applications

Analysis of Time Series Models on National Scale

The time series models are stored in a database. The time series of about 34,000 piezometers satisfy the selection criteria for time series modeling. A large part of them can be classified as “potentially useful” or better. The results of these models have been analyzed and spatial patterns of the classification and specific model output have been created.

For reasons of visualization, values have been averaged within a 5 km grid. Figure 3 shows the relative success rate of the time series modeling. The left pane of Figure 3 shows the number of accepted models (MODOK = 1, model is potentially useful or better) per grid cell divided by the number of time series which meet the requirements for creating a model (TSOK = 1). In the right pane of Figure 3 the average groundwater depth is depicted. As can be seen, some areas have a low fraction of accepted models, most notably the Southeastern (Zuid-Limburg) and central (Veluwe) parts of the Netherlands. These are the areas with relatively large groundwater depths. This suggests that the current setup with a linear model fitted to 8 years of head data is less suitable in areas with a large unsaturated zone. The other areas (with mostly shallow groundwater depths) do not show a distinct pattern for model acceptance.

Regional Analysis of Aquitard

The time series models have been used to get insight in the functioning of an aquitard in the area of the city of Zwolle. This insight was required because of spatial planning problems. One question was whether groundwater resources are protected by the presence of an aquitard above an elevation of approximately 75 m below sea level. Figure 4 shows a cross section along three multi piezometer wells in the hydrogeological model REGIS II (<http://www.dinoloket.nl>).

In the REGIS II model, there is an aquitard on the East side (KRTWk1). Its influence on the groundwater flow is clear from a head difference between the heads measured above and below this aquitard. The complex unit DTc consists of—mostly sandy—sediments that have been pushed by a glacier in the Saalien ice age. In well B21G0491 there is no significant head difference between piezometers above and below this layer, but the response functions for precipitation are quite different, indicating that the layer does have some confining properties at this location. So, the clay in the borelog at a depth of 100 m (see Figure 5) does have

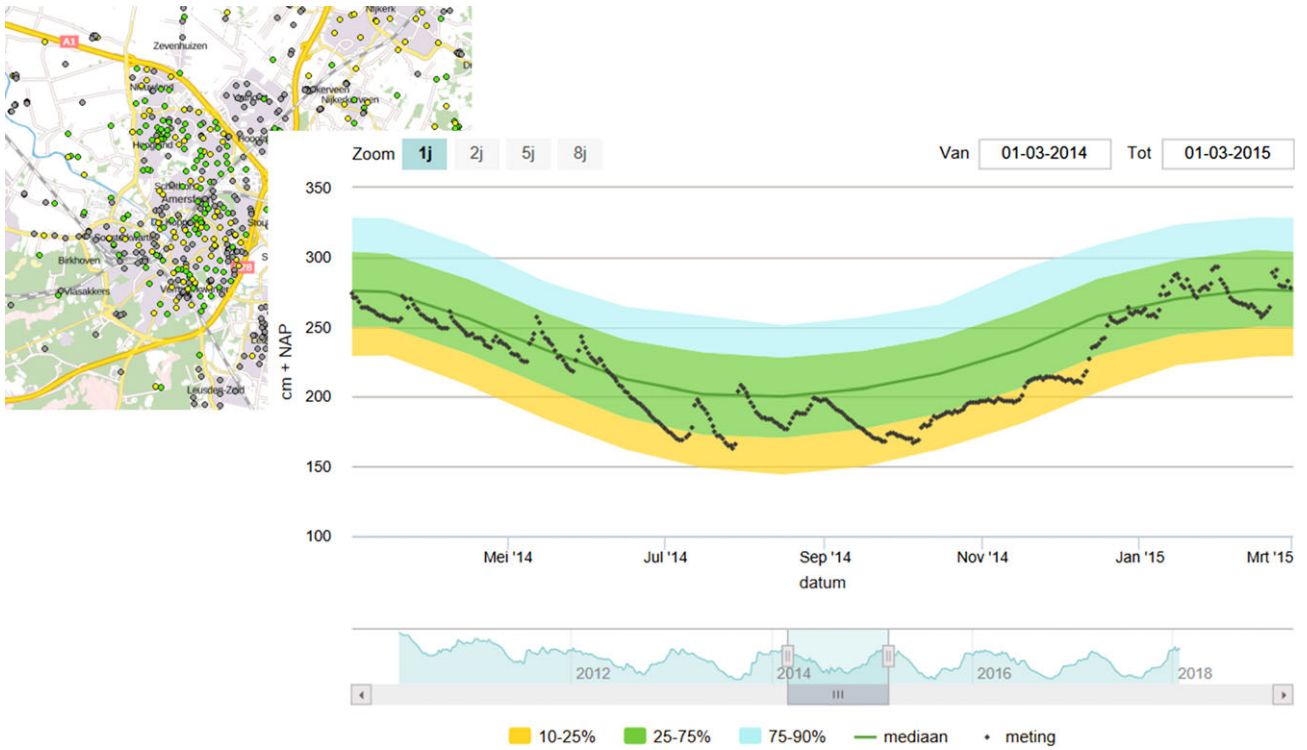


Figure 2. Visualization of the regime curve with average high (GHG), average low (GLG), and average spring (GVG) level of the piezometric head in the web interface with in the background the map with quality of time series models (gray = insufficient; yellow = potentially useful; green = decent or good).

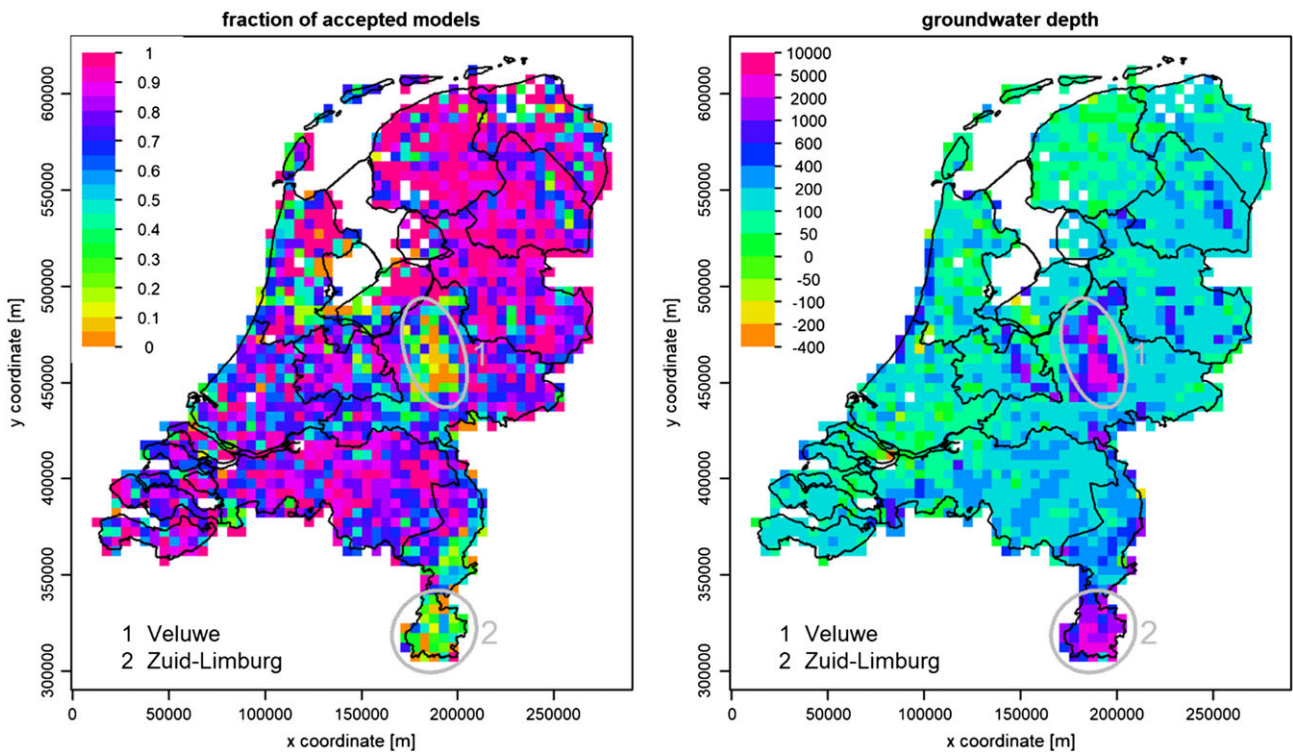


Figure 3. Fraction of accepted models (left), and depth of the groundwater heads in cm below surface (right).

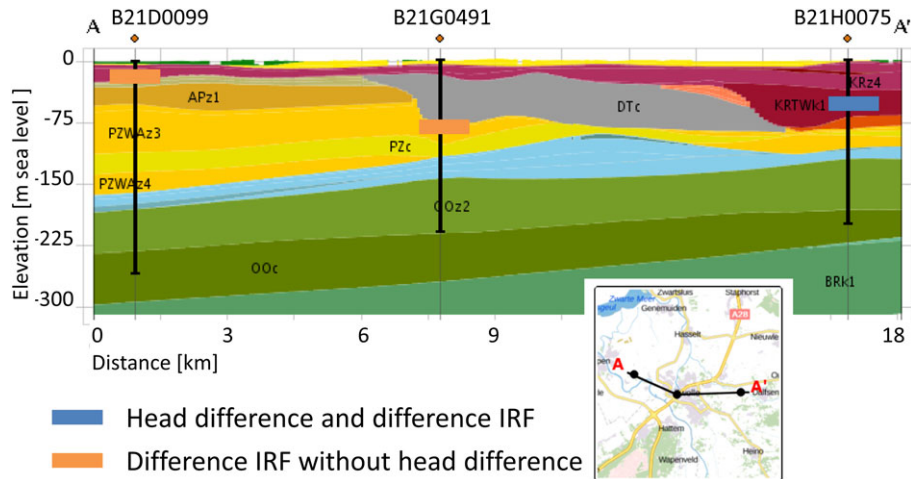


Figure 4. Cross section around Zwolle showing locations of observation wells and the hydrogeological model REGIS II with geological formations (capitals) and aquifer (z), aquitard (k), and complex (c) hydrogeological units.

enough lateral extend to create a difference in precipitation response.

In well B21D0099, no regional aquitard was modeled at an elevation of 25 m below sea level (see Figure 4), but precipitation response is different for the piezometers below and above this level. Again, the borehole description (see Figure 5) does show some clay. It was not considered important enough to map it as part of a regional aquitard (due to the absence of a head difference). However, the extent is large enough to cause a change in the precipitation impulse response function of the piezometric head across the layer.

Selection of Piezometers for Checking Groundwater Models

Another application of time series models is the selection of piezometers for (preliminary) checking of a groundwater model (Zaadnoordijk and Bakker 2013). Two aspects can be considered: the quality of the time series model and the consistency with neighboring piezometers.

A good quality of the time series model indicates that the piezometric head can be explained by precipitation and evaporation. A groundwater model should definitely be able to reproduce these measurements, assuming that these stresses are included in the boundary conditions of the groundwater model. There is no “excuse” for the groundwater model, that the measurements are strongly influenced by a stress that is not included.

The consistency of the transfer functions in the time series model with neighboring piezometers shows how variable the behavior of the groundwater system is spatially. It may be necessary to account for the resolution of the groundwater in the comparison with the measured heads when the transfer functions vary at a small scale. Otherwise systematic differences may be expected in the comparison of model output with heads from the piezometers.

Time series models also can be used to focus the calibration of a groundwater model. If the effect of

a specific stress is important for the purpose of the groundwater model and if it also is possible to determine the effect of this stress in transfer noise models of the time series from the piezometric heads, then the groundwater model can be calibrated using the impulse response functions from the time series models. This way, the calibration specifically addresses the purpose of the groundwater model and a short time period can be used for the calibration run, since only the effect to an impulse of this stress has to be simulated and not a real period in which the stress has sufficient variation different from other stresses (Zaadnoordijk and Bakker 2013).

Creation of Consistent Statistics for Piezometers

The time series of the piezometers in the Dutch national database vary in measurement period, frequency, and regularity. This makes that the average of the measurements per piezometer and other statistics cannot be used directly for analyses. For the piezometers with a good enough transfer function-noise model, simulations can be carried out to produce simulated heads for one and the same period and frequency for all piezometers. The results can be processed to provide statistics that are consistent. This way it is possible to consistent values for the previously mentioned statistics (average high, average low, and average spring groundwater level) which are commonly used in groundwater management in the Netherlands (Van Heesen 1970; Finke et al. 2004).

Discussion

Performing time series analysis on a large number of head time series requires automation of the time series modeling and evaluation process. It is impossible to judge the adequacy of these criteria without considering the models in more detail, and more importantly, having a purpose in mind. So, it remains up to the user to evaluate a particular model of interest in the web interface, before using the results. Feedback from users and insights gained

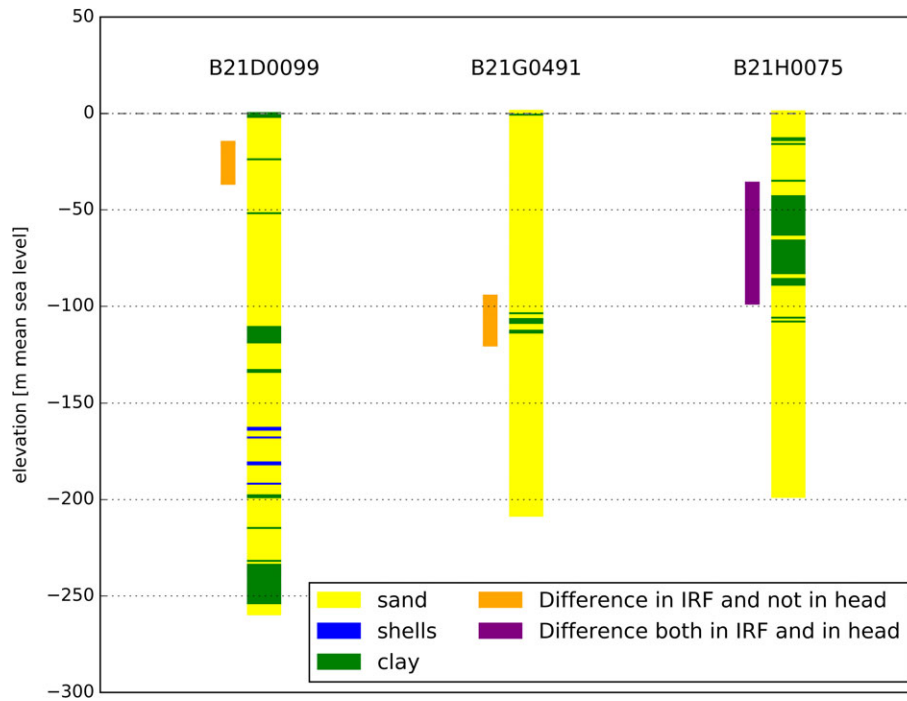


Figure 5. Descriptions of boreholes in cross section.

from analyses like the ones presented in the Results section will help to improve our general purpose criteria in the future. Scientific challenges remain with respect to the validity of the derived impulse response functions. In the current setup, precipitation and evaporation are separated from other influences. It is not clear how to account for the fact that the other influences have an unknown character in time when the reliability of the precipitation and evaporation impulse response functions is determined. Non-linearity of the groundwater systems makes that the influence of precipitation and evaporation cannot be represented by a single impulse response function. It still is unclear how to account for this aspect in the validity of the derived impulse response functions.

The choice has been made to model time series for the last 8 years of observations only. This ensures the time series models reflect the current response of the groundwater system and the time series modeling is not hampered by changes in the system. On the other hand, changes/trends in the system can be detected by evaluating other time periods and comparing the results. A disadvantage of using only 8 years of data is that the model uncertainty increases due to the limited amount of data. This especially makes a difference in areas with slower response to precipitation. Therefore, it may explain the lower success rate of the time series modeling in areas with a large unsaturated zone (see Figure 3).

Expanding the model with the input of river and abstraction stresses (e.g., Obergefell et al. 2013; Shapoori et al. 2015) could increase the number of acceptable models. However, in an automated modeling procedure, it may give higher risks of over-parametrization, as it is not clear how to decide automatically which stresses to

include for a particular piezometer. Besides, there is a data problem: water levels for smaller water bodies and abstraction data are not readily available for the required periods at an adequate frequency. An alternative approach could be to evaluate nearby piezometers together using dynamic factor analysis (Berendrecht and van Geer 2016). In this approach, the residuals of the time series models with precipitation and evaporation are converted into one or more common dynamic factors and one specific factor for each piezometer. The time series of the common factors may give information on other regional influences while the time series of the specific factors may help to detect data problems.

The impulse response functions of the time series models include information about the unit step response and the response time. The spatial patterns of these quantities provide useful information about the groundwater systems. The total reaction is given by the area underneath the response function, which is equivalent to the unit step response. The delay time of the reaction of the groundwater head to the input signal can be determined from the shape of the impulse response function. In Figure 4 the unit step response and the median response time to precipitation are displayed as averages per grid cell of 5 km by 5 km. The distributions of high and low values are comparable with large total response and long response times in areas with relatively large depths of the groundwater table. Areas with much surface water show a small unit step response and short response times.

Figure 6 shows average values per 5 km by 5 km grid cell. The results of the times series models also can be used to interpolate the influence of precipitation spatially. It even is possible to include hydrological

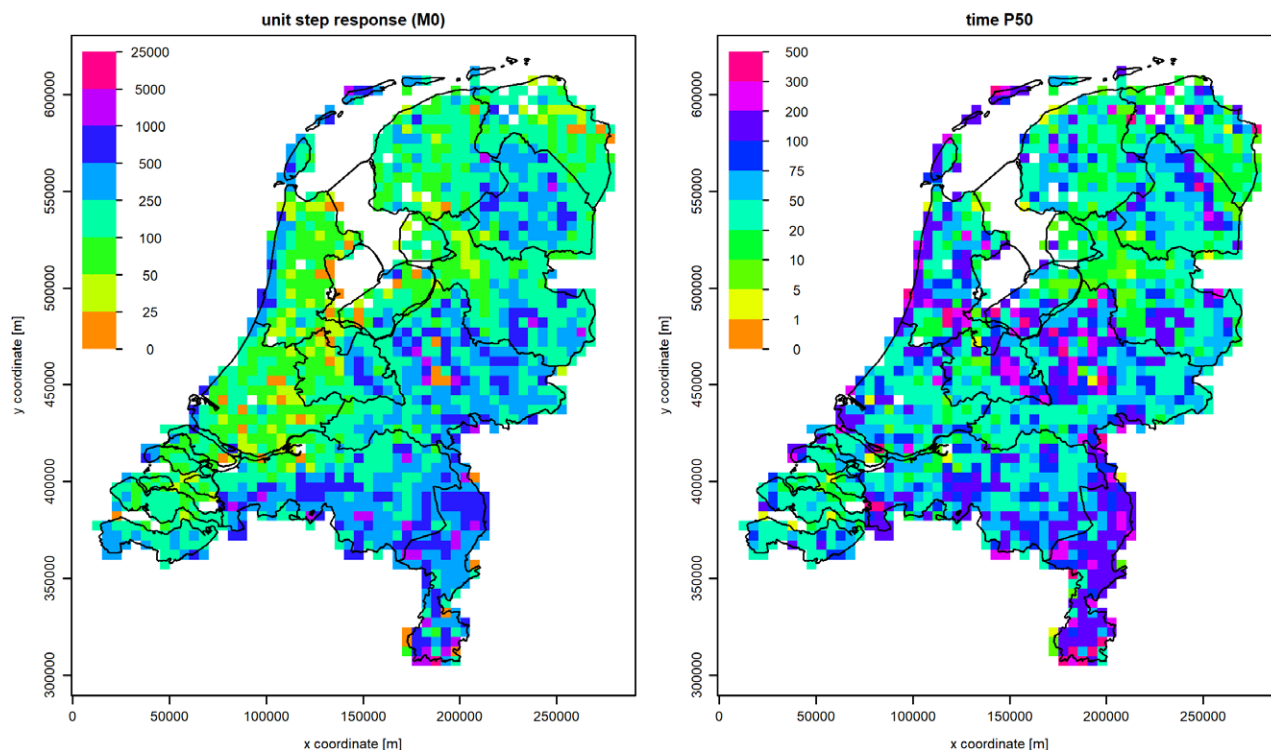


Figure 6. Characteristics of the response functions for precipitation: unit step response in days (m/[m/d], left) and median response time in days (right).

boundaries in the interpolation, because the moments of the response functions fulfill differential equations that are similar to those of groundwater flow (Bakker et al. 2007; Zaadnoordijk and Bakker 2013) This can be used as an alternative for a physically based groundwater model to predict groundwater levels with much less computation time (Van Loon and Zaadnoordijk 2015).

The possibility to detect aquitard in areas without a vertical gradient gives the availability of the large number of time series models particular value for geohydrological characterization. This was illustrated by the investigation of an aquitard near Zwolle. It is hard to detect the existence of an aquitard in the calibration of a groundwater model if no systematic head difference across the layer exists. Both the target function for automatic parameter optimization and inspection of differences between model output and measurements require aggregation of the deviations of individual measurements, which make it hard to detect the fluctuations of head difference which are needed to adjust the aquitard extent.

In addition to delineation of aquitards, it may be possible to use the transfer function-noise models for improvement of the hydraulic parametrization following Bakker et al. (2008) and Obergfell et al. (2013) who linked time series models to subsurface properties. This could help to improve the hydraulic conductivity in the models REGIS II and GeoTOP that the TNO Geological Survey of the Netherlands maintains (<https://www.dinoloket.nl/en/subsurface-models>).

With the application of the time series models for selection of calibration targets of a physically based

groundwater model, we do not want to suggest that piezometers with a bad time series model should not be used for calibration of such a model. We only want to point out that there are fewer reasons to accept that the model does not represent measurements from piezometers with a good time series models. Then, it is unlikely that there is a large influence from a (local) stress not included in the model or that the measurements contain large errors.

The automated procedures for creating and judging time series models will aid quality control of the piezometric data. Also they can be used to detect changes in the groundwater system and separate natural and anthropogenic changes which is importance for policy making related to groundwater resources and planning (groundwater dependent) land use.

Conclusions

The response of the groundwater head to precipitation and evaporation has been determined using automated transfer function-noise time series modeling for piezometers in the Dutch national database with piezometric heads. The resulting time series models can be accessed through a public web site depending on the quality of the model. The chosen model quality criteria seem to be acceptable for the automated model assessment.

Next to the additional information created per piezometer, the large collection of time series models provides information of the groundwater system which can be used for various purposes.

References

- Bakker, M., K. Maas, and J.R. Von Asmuth. 2008. Calibration of transient groundwater models using time series analysis and moment matching. *Water Resources Research* 44: W04420. <https://doi.org/10.1029/2007wr006239>
- Bakker, M., K. Maas, F. Schaars, and J.R. Von Asmuth. 2007. Analytic modeling of groundwater dynamics with an approximate impulse response function for areal recharge. *Advances in Water Resources* 30: 493–504.
- Berendrecht, W.L., and F.C. van Geer. 2016. A dynamic factor modeling framework for analyzing multiple groundwater head series simultaneously. *Journal of Hydrology* 536: 50–60.
- Besbes, M., and G. de Marsily. 1984. From infiltration to recharge: Use of a parametric transfer function. *Journal of Hydrology* 74: 271–293.
- De Bruin, H.A.R., and W.N. Lablans. 1998. Reference crop evapotranspiration determined with a modified Makkink equation. *Hydrological Processes* 12, no. 7: 1053–1062. [https://doi.org/10.1002/\(SICI\)1099-1085\(19980615\)12:7<1053::AID-HYP639>3.0.CO;2-E](https://doi.org/10.1002/(SICI)1099-1085(19980615)12:7<1053::AID-HYP639>3.0.CO;2-E)
- Finke, P.A., D.J. Brus, M.F.P. Bierkens, T. Hoogland, M. Knotters, and F. de Vries. 2004. Mapping groundwater dynamics using multiple sources of exhaustive high resolution data. *Geoderma* 123: 23–39. <https://doi.org/10.1016/j.geoderma.2004.01.025>
- Gehrels, J.C., F.C. Van Geer, and J.J. De Vries. 1994. Decomposition of groundwater level fluctuations using transfer modelling in an area with shallow to deep unsaturated zones. *Journal of Hydrology* 157: 105–138.
- Kim, S.J., Y. Hyun, and K.K. Lee. 2005. Time series modeling for evaluation of groundwater discharge rates into an urban subway system. *Geosciences Journal* 9: 15–22.
- Manziona, R., M. Knotters, G. Heuvelink, J. Von Asmuth, and G. Camara. 2010. Transfer function-noise modeling and spatial interpolation to evaluate the risk of extreme (shallow) water-table levels in the Brazilian Cerrados. *Hydrogeology Journal* 18, no. 8: 1927–1937.
- Obergefell, C., M. Bakker, W.J. Zaadnoordijk, and K. Maas. 2013. Deriving hydrogeological parameters through time series analysis of groundwater head fluctuations around well fields. *Hydrogeology Journal* 21: 987–999. <https://doi.org/10.1007/s10040-013-0973-4>
- Peterson, T.J., and A.W. Western. 2014. Nonlinear time-series modeling of unconfined groundwater head. *Water Resources Research* 50: 8330–8355. <https://doi.org/10.1002/2013WR014800>
- Russo Tess, Upanu Lall, Hui Wen, and Mary Williams (2014) Assessment of trends in groundwater levels across the United States, Columbia Water Center White Paper. water.columbia.edu (accessed November 14, 2017).
- Schoups, G., and J.A. Vrugt. 2010. A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. *Water Resources Research* 46, no. 10: W10531. <https://doi.org/10.1029/2009WR008933>
- Shapoori, V., T.J. Peterson, W. Western, and J.F. Costelloe. 2015. Top-down groundwater hydrograph time-series modeling for climate pumping decomposition. *Hydrogeology Journal* 23: 819–836. <https://doi.org/10.1007/s10040-014-1223-0>
- Stoffer, D.S., and C.M.C. Toloi. 1992. A note on the Ljung-box-pierce portmanteau statistic with missing data. *Statistics & Probability Letters* 13: 391–396.
- Thyer, M., B. Renard, D. Kavetski, G. Kuczera, S.W. Franks, and S. Srikanthan. 2009. Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using Bayesian total error analysis. *Water Resources Research* 45: W00B14. <https://doi.org/10.1029/2008/WR006825>
- Van Heesen, H.C. 1970. Presentation of the seasonal fluctuation of the water table on soil maps. *Geoderma* 4, no. 3: 257–278.
- Van Loon, A., and W.J. Zaadnoordijk. 2015. Vlakdekkende tijdreeksanalyse: Een data-gedreven methode voor het projecteren van grondwaterstand reeksen (Spatial time series analysis: a data driven method for projecting groundwater level series; in Dutch). *Stromingen* 23, no. 3: 37–51.
- Von Asmuth, J.R., M.F.P. Bierkens, and K. Maas. 2002. Transfer function-noise modeling in continuous time using predefined impulse response functions. *Water Resources Research* 38, no. 12: 23-1–23-12.
- Zaadnoordijk, W.J., and M. Bakker. 2013. *Application of Spatial Time-Series Analysis to Determine Calibration Targets for Transient Groundwater Models, MODFLOW & More 2013*. Golden, Colorado: IGWMC.