



## Deliverable 3.5

# 3D visualization of cross-border patterns of groundwater depletion

Authors and affiliation:

**Willem Jan Zaadnoordijk, Jelle Buma, Ronald Vernes**  
TNO Geological Survey of the Netherlands

**Koen Beerten, Bart Rogiers**  
SCK•CEN

**Cis Slenter**  
VMM

**Renate Jaritz**  
Ertverband

**Bernd Linder**  
Geologischen Dienst Nordrhein-Westfalen

**Hans Peter Broers**  
TNO Geological Survey of the Netherlands

E-mail of lead author:

**willem\_jan.zaadnoordijk@tno.nl**

Version: final

This report is part of a project that has received funding by the European Union's Horizon 2020 research and innovation programme under grant agreement number 731166.



Deliverable Data		
<b>Deliverable number</b>	D3.5	
<b>Dissemination level</b>	Public	
<b>Deliverable name</b>	3D visualization cross-border patterns of groundwater depletion	
<b>Work package</b>	WP3, H3O-PLUS	
<b>Lead WP/Deliverable beneficiary</b>	VMM	
Deliverable status		
<b>Submitted (Author(s))</b>	16/07/2021	Willem-Jan Zaadnoordijk
<b>Verified (WP leader)</b>	23/07/2021	Cis Slenter
<b>Approved (Coordinator)</b>	26/07/2021	Hans Peter Broers



## TABLE OF CONTENTS

1	INTRODUCTION.....	3
2	DATA.....	5
3	METHOD.....	7
	3.1 Head analysis per country .....	7
	3.2 Heads for entire H3O-PLUS area .....	30
4	RESULTS.....	33
	4.1 Analysis per country .....	33
	4.2 Analysis for entire H3O-PLUS area.....	53
5	CONCLUSIONS .....	66
6	REFERENCES.....	67
	APPENDIX A – DATABASE.....	71
	APPENDIX B – EVAPOTRANSPIRATION (ET) VS TEMPERATURE (T) .....	72
	APPENDIX C – IMPULSE-RESPONSE FUNCTION MODEL ARCHITECTURES .....	84

# 1 INTRODUCTION

The cross-border demonstration project H3O-PLUS aims to set a new standard for harmonization across borders, not only for hydrostratigraphy, but also for hydrological data such as groundwater heads and groundwater quality.

H3O-PLUS, WP3 of Resource, aims to be an advanced demonstration of a transboundary assessment of groundwater resources. It is 'advanced' in the sense that it builds on and extends previous work, trying to make it more useful for groundwater policy and management and for subsurface spatial planning. A 3D hydrogeological model has been developed in a series of so called 'H3O' projects in the transboundary region around the Roer Valley Graben, comprising parts of Germany, the Netherlands and Belgium. The model contains 3D maps of the top, base and thickness of aquifers and aquitards (see Figure 1.1). H3O-PLUS aims to add attribute data to these maps to facilitate the use of the maps in decision making processes. Note that the project does not aim to produce new maps or spatial delineations. The objective is to characterize units on existing maps and hence support the interpretation and use of those existing maps.

The overall study area coincides with the study areas of previous H3O projects (Figure 1.1). Vertically, the study is limited to the clastic (hydro)geological layers of Cenozoic age or younger. This coincides with the vertical scope of the recently developed transboundary 3D (hydro)geological models of the H3O projects. The base of the models is thus located at the top of the Chalk aquifer (Formation of Houthem or Maastricht) or the top of the Carboniferous deposits.

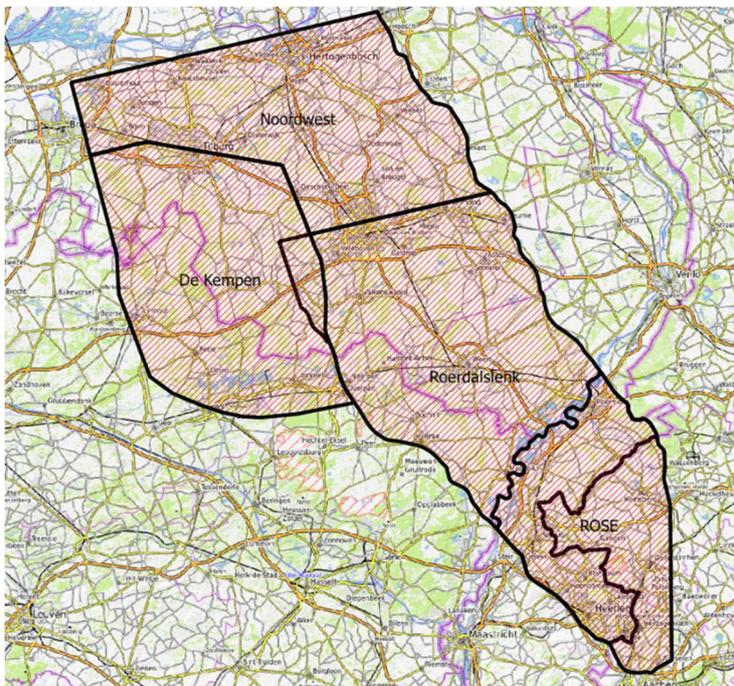


Figure 1.1: Study areas of previous H3O projects (note the label “Roerdalslenk” can be translated to “Roer Valley Graben”)



This report describes the current practises of evaluation of the groundwater quantity monitoring in the three countries for groundwater management. It also covers the set-up of a database of groundwater head data and meteorological data for the H3O area covering parts of Belgium, Germany, and the Netherlands. The data is analysed, and the results are visualised in maps and transects. Groundwater head data within the provinces Noord-Brabant and Limburg from the Dutch national subsurface data repository (<https://www.DINOloket.nl>), Flemish data from VMM and SCK, and German data within North Rhine-Westphalia from the Erftverband have been added, together with precipitation, evaporation, and temperature data from the Dutch, Belgian, and German Meteorological Institutes. The report represents the first transboundary visualisation of groundwater heads and cross-border patterns of groundwater depletion for the study region based on the respective monitoring networks in the participating countries. These visualisations are now available through a dedicated transboundary GeoERA Groundwater web viewer. The results of the current report will be input for the next deliverable (D3.6) where we will explore the current protection and management strategies for shallow resources and deep paleo resources in the region. For this aim, the consortium cooperates with the regional stakeholders involved in the project, using the newly developed information and visualization system to help groundwater managers to harmonize groundwater strategies. As such, the work has been supported in part by the Dutch provinces of Noord-Brabant and Limburg, as well as the water companies Brabant Water and WML.



## 2 DATA

In addition to a presentation of the national monitoring practices and connected results, a database has been created with groundwater head timeseries and meteorological data from Belgium, Germany and the Netherlands. The database has been analysed uniformly for the project area using various methods and the results are made available in a web-based viewer. The methods include trend analysis and time-series modelling using a transfer-noise approach linking the groundwater fluctuation to meteorological variation.

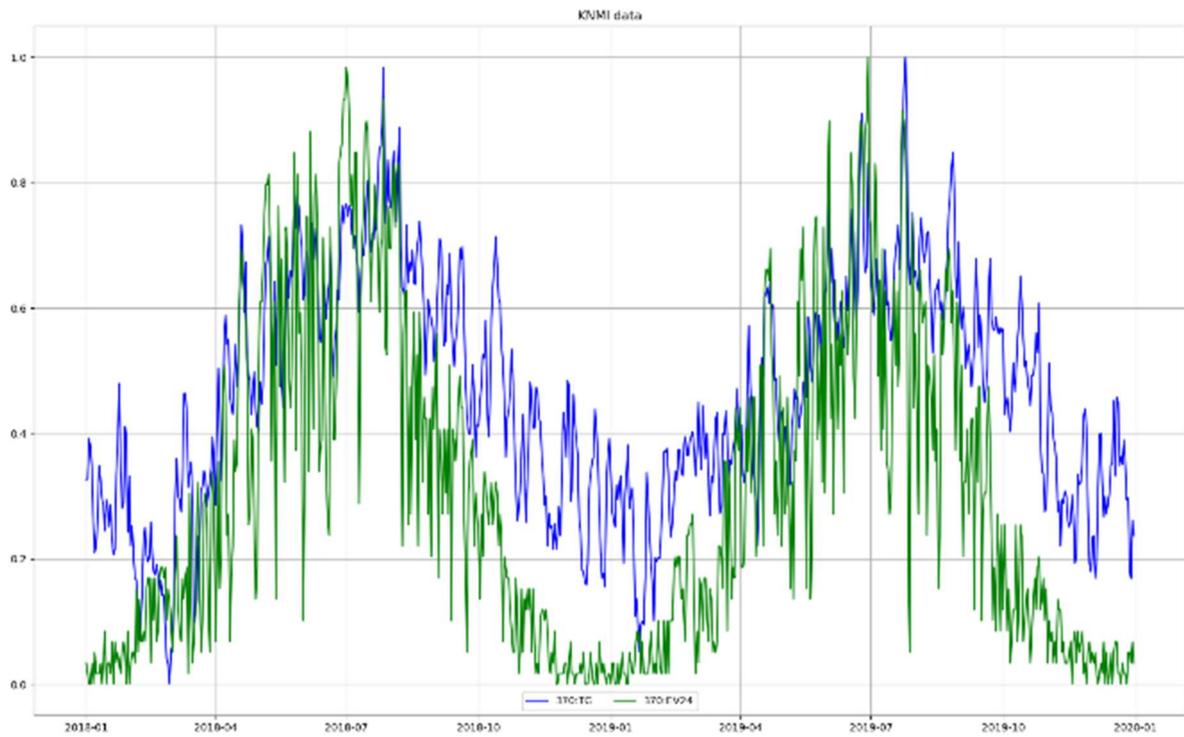
The following criteria were used for the selection of the timeseries of groundwater heads:

- At least 15 years with observations in  $\geq 6$  months, within the period 1994 through 2018. Analysis further back into history is not desirable, because of an increased probability of changes in the hydrological system that may not be traceable;
- Selection of piezometers in subsurface above Krijt/Kreide, in the project area with a buffer of 5 km around it. This covers the depths and area relevant for the studied groundwater resources;
- The date of the last measurement is on 1/1/2017 or later. Thus, limiting the selection to piezometers that were in use at the end of the data period.

The meteorological data consists of timeseries of precipitation, evapotranspiration, and temperature:

- One value per day (daily totals of precipitation and evapotranspiration, and daily averages of temperature);
- Covering period 1/1/1993 through 31/12/2018 (end date is identical to end date of groundwater head data period).

For Belgium and Netherlands, Makkink evapotranspiration data were used. For Germany no evaporation data was available. Therefore, temperature has been used as a proxy, because of the correlation between the variation of temperature and evaporation (Figure 2.1).



*Figure 2.1 Relative values of average daily temperature (TG, blue line) and Makkink evaporation (EV24, green line) for the Dutch KNMI station of Eindhoven (370).*



## 3 METHOD

### 3.1 Head analysis per country

#### 3.1.1 Belgium

The Belgian part of the H3O-Plus area is located in the Flemish part of Belgium, Flanders.

##### 3.1.1.1 Organisation

The Flanders Environmental Agency (Vlaamse Milieumaatschappij, VMM) is responsible for the groundwater monitoring in the Flemish part of Belgium. For the status and trend monitoring they use two monitoring networks: the primary groundwater monitoring network and the phreatic groundwater monitoring network. Beside these there are other groundwater monitoring networks for example from nature organisations and for the follow up of bigger groundwater abstractions and projects to monitor the impact on the groundwater. The groundwater monitoring networks can be found on [dov.vlaanderen.be](http://dov.vlaanderen.be).

The primary groundwater monitoring network consists of single and multilevel monitoring wells where the groundwater level is measured manually each month. The screens are evenly distributed over the various aquifers, as far as possible situated outside the perimeter of (direct) anthropogenic influence. There are 74 wells with a total of 167 screens. The depth ranges from 2 to 570 meters. The “GroundWater head Indicator (GWI)” monitoring network (for monitoring of drought events) is part of the primary groundwater monitoring network. The GWI monitoring points are screens which are used for the trend analysis of measured and predicted phreatic groundwater heads. These groundwater heads provide the most reliable picture of recent climate variations, that are tested against a relatively long history of terrain measurements.

The configuration of the multilevel wells of the phreatic groundwater monitoring network is based on the behaviour of nitrates. Where the top filter is located directly below the water table and the second and third filter are located just above and just below the oxidation-reduction zone, respectively. The groundwater heads are measured here manually twice a year.

The location of the monitoring wells in the Belgian H3O-Plus area is shown in Figure 3.1.

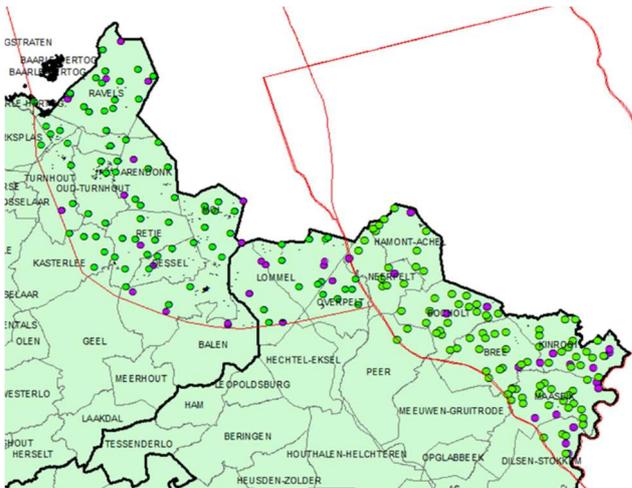


Figure 3.1: Primary monitoring wells (purple dots) and the phreatic monitoring wells (green dots) of VMM in the Belgian part of the H30-PLUS area.

### 3.1.1.2 Analysis method

#### Groundwater time series

The groundwater time series of the VMM can be analyzed and questioned at [dov.vlaanderen.be](http://dov.vlaanderen.be).

#### Trend calculation

VMM uses different approaches for trend detection for phreatic and confined groundwater.

For the trend calculations of the phreatic groundwater, the monitoring data of the primary and phreatic groundwater monitoring network are used. The calculations are performed using the “Trendanalist 6” program (<https://www.amo-nl.com/software/trendanalist/>).

Trendanalist determines a monotonic linear trend on the measurement series and chooses a suitable significance test based on the characteristics of the measurement series: this is a linear regression test and its extensions or a Mann-Kendall test and its extensions. A test from the first group is applicable if the values of the measurement series have a normal probability distribution. If this is not the case, a test from the second group applies. For both groups, the extensions are intended to also take into account seasonal effects and/or autocorrelation. Trendanalist tests whether there is a monotonous trend in the data. The level of significance was set to 0.05.

Trend reversals and establishing the starting points for trend reversal cannot be determined using the foregoing method.

This methodology is used as part of the status and trend assessment of the groundwater bodies for the River Basin Management Plans (WFD). The methodology does not take into account or does not distinguish between natural and anthropogenic impact. This is justifiable because the situation is effectively influenced by the combination of natural and anthropogenic effects and a groundwater policy must take into account both and respond to them. The methodology cannot determine for the contribution of individual

influences to the trend. This is necessary to be able to conduct a responsible multi-track policy for groundwater and to determine the effectiveness of measures.

### Trends on the SWAP-series of the indicator points

In order to gain some insight into the proportion of natural and anthropogenic variation in the trends of the phreatic measurement series, the GWI monitoring network is used. Spread across Flanders, this monitoring network comprises a selection of approximately 154 screens from the primary groundwater monitoring network. The screens are selected in such a way that they are phreatic and that the groundwater levels are influenced as little as possible by water extraction, drainage or other human interventions. In addition to the monthly measurements, a SWAP model (Soil, Water, Atmosphere and Plant model, a physically based 1-dimensional model of the unsaturated zone created by the SWAP software. <https://www.swap.alterra.nl>) was developed for each screen to simulate the groundwater heads in this monitoring point (Heuvelmans et al., 2011). The Indicator monitoring network is usually used to draw up the groundwater head indicator ([Grondwaterstandindicator | DOV \(vlaanderen.be\)](http://Grondwaterstandindicator | DOV (vlaanderen.be))) which, among other things, monitors drought.

SWAP simulates the hydraulic head for shallow, phreatic screens (<50m -mv) and takes into account, among other things, soil properties, the observed daily precipitation and evaporation etc., but does not take into account anthropogenic influences (such as groundwater extraction). Thus, trends on the simulated hydraulic heads are a measure of the variation in climatic conditions. These simulated groundwater head series can then be compared with the measurements (see Figure 3.2). The difference between the two cannot be explained by SWAP and is therefore an indication of the anthropogenic influence, but also includes the model uncertainties (Kroes et al., 2008).

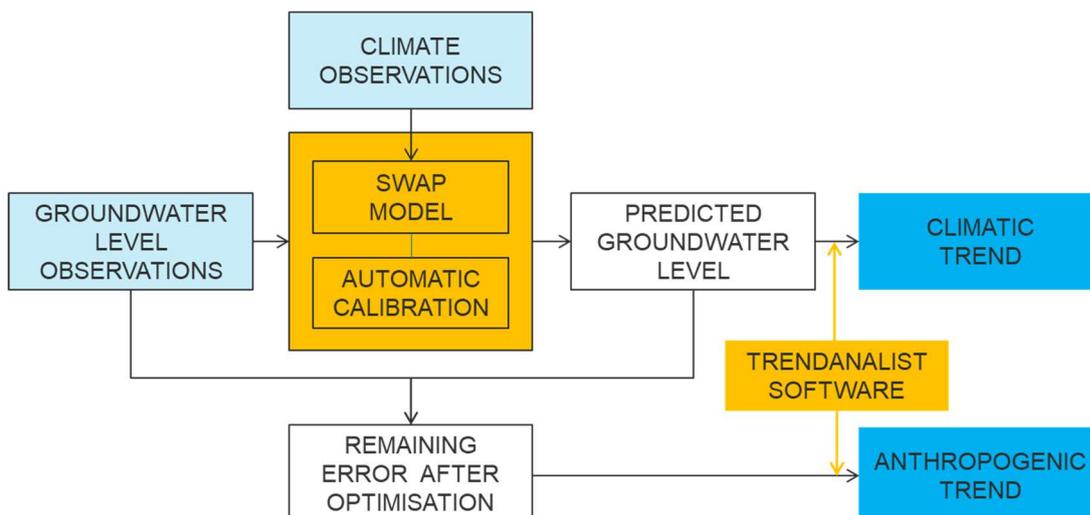


Figure 3.2 Trend determination method of VMM for phreatic groundwater heads.

Trends are also calculated on the simulated levels and the residuals calculated as a difference with the site measurements for various indicator points, using the statistical program Trendanalist.

Any statement regarding the share of climatic versus anthropogenic influence on the particular trend, however, can only be made at screen level and cannot be interpreted globally.

For confined groundwater a statistical approach is used. The confined head time series are pre-processed to fill gaps and remove outliers in the time series. The method for gap filling depends on the data availability. If a time series is more than 75% complete, the gaps are filled by a time series model (SARIMAX) based on the detrended series. For series with 15-75% of data, the gaps are filled based on correlations between head time series in a groundwater body and meteorological variables.

Next, LOESS is used for an initial assessment of the trends. Regression is carried out per segment of each series. Finally, the trend changes are refined (see Figure 3.3).

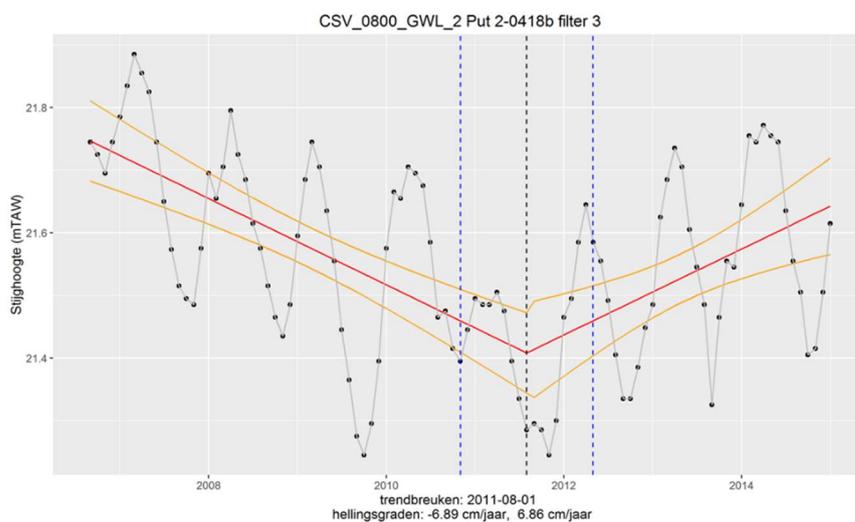


Figure 3.3 Example of trend determination for confined groundwater heads

### 3.1.1.3 Visualization method

The trends within a waterbody can be summarized in a single block diagram (Figure 3.4).

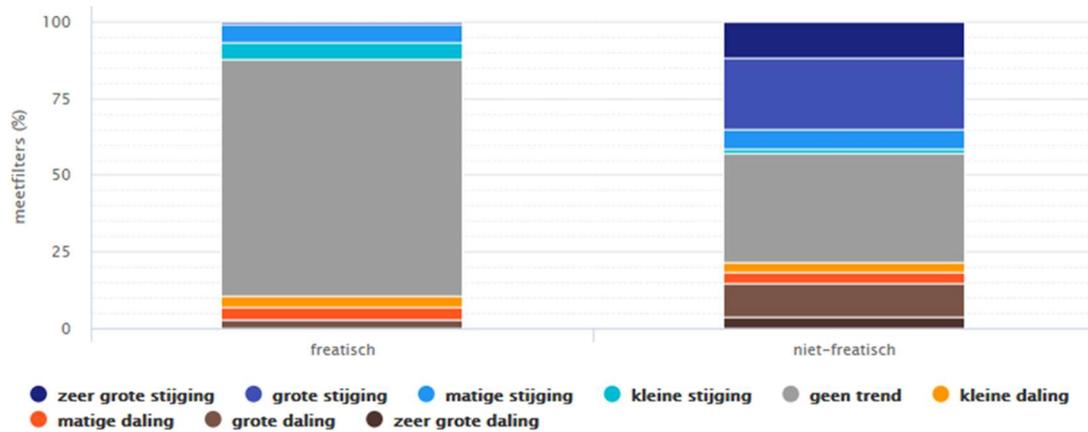


Figure 3.4 Visualization of trends in a phreatic (left) and confined (right) water body with trends ranging from very large increase (dark blue) to a very large decline (dark brown).

In addition to a single trend, VMM also visualizes the variation of the groundwater head in monitoring wells accounting for changes in the trends (Figure 3.5).

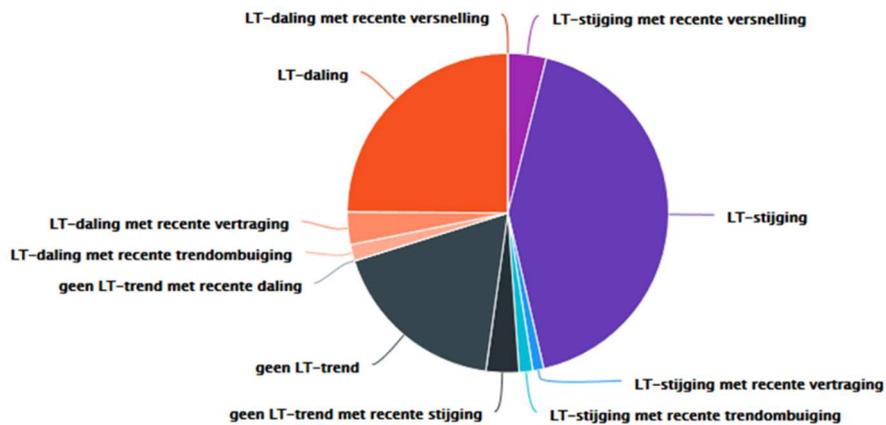


Figure 3.5 Visualization of trends in confined groundwater with an indication of changes of the rates ('versnelling'=acceleration, 'vertraging'=deceleration, 'trendombuiging'= trend reversal, 'stijging'=increase, 'daling'=decline).

The spatial distribution of the trends is also visualized (Figure 3.6).

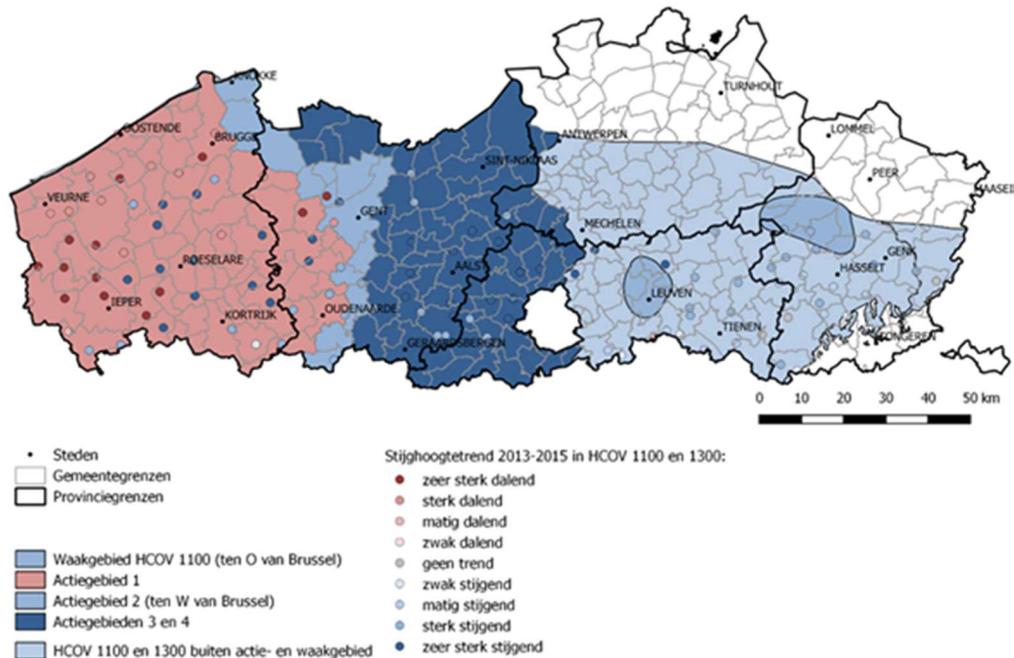


Figure 3.6 Visualization of spatial distribution of trends ranging from very large increase (dark blue points) to a very large decline (dark brown points). The background color of the management areas (in this example the action and watch areas of the recovery programs that are posed for groundwater bodies in poor quantitative status) shows the median of the trends of the filters located in the respective action area.

#### 3.1.1.4 Impulse-response function modelling

While the main experiences of the SCK CEN with impulse-response function (IRF) type of modelling were mainly with the RRAWFLOW model (Long 2015), with a custom R (R Core Team 2021) wrapper around for the optimisation of the model parameters (Rogiers et al. 2016, 2018; Lu et al. in prep), for this work we wanted to use a more flexible toolset which would make testing different IRF model architectures more straightforward. Hence, we resorted to an embedded probabilistic programming language (Golding 2019) and implemented the IRF models there. For now, because of time constraints, just aiming at optimisation, but this does enable proper probabilistic Bayesian inversion, which could be explored in future.

Although RRAWFLOW uses already a rather complex architecture (non-linear recharge estimation with options for time-variant and non-parametric IRFs), we decided to look here at a larger range, of simpler architectures, to see what the effects on the estimated trends would be. As we saw previously that the time-variant IRF option in RRAWFLOW was useful, this is tested here as well, by allowing small fluctuations of the IRF parameters on a (periodic) monthly basis. The parametric IRF used here is the typical gamma distribution function with three parameters, as used by e.g. Long (2015) and



Zaadnoordijk et al. (2019), and all three are allowed to vary (albeit in a very limited way, as we use regularizing Gaussian priors for the deviations).

Furthermore, RRAWFLOW does not include a particular approach for addressing temporal correlation of the model residuals, while this is commonly used elsewhere in literature (e.g. Von Asmuth et al. 2002, 2008; Von Asmuth & Bierkens 2005; Obergfell et al. 2019; Zaadnoordijk et al. 2019; Collenteur et al. 2021). Additionally, one of the goals here is to quantify the anthropogenic, or maybe better the “unexplained” trends in the groundwater level time series. Therefore, it was decided to test the use of Gaussian processes (Roberts 2013) here, using Golding (2020), to both explicitly model the unexplained part of the time series, as well as address the temporal correlation of the model residuals.

Additionally, we were working on process models for groundwater flow in the unsaturated and saturated zones for different catchments, in the framework of the Future floodplains project (<https://www.futurefloodplains.be/>). Some of the recharge calculations are done there for an area not too far from the H3O-plus project area. Hence it was also interesting to try and see if these results could be used for the input to these IRF type of models, next to the more traditional use of meteorological time series (i.e. precipitation and temperature in case of RRAWFLOW, or precipitation and evapotranspiration in this work), which does not require a separate process modelling exercise.

The general approach here is thus to look at all possible combinations of the above (i.e. inputs to the model, time-variance (or not) of the IRF and the inclusion (or not) of a Gaussian process), all implemented in the embedded probabilistic programming language, and applied to the set of 21 indicator wells of the VMM (see Figure 3.7). The goal of this exercise is purely exploratory here; by no means we are trying to make confirmatory statements on the different investigated IRF model architectures. We are just trying to get a feeling of the effects of different input time series and model architectures on the estimated trends.

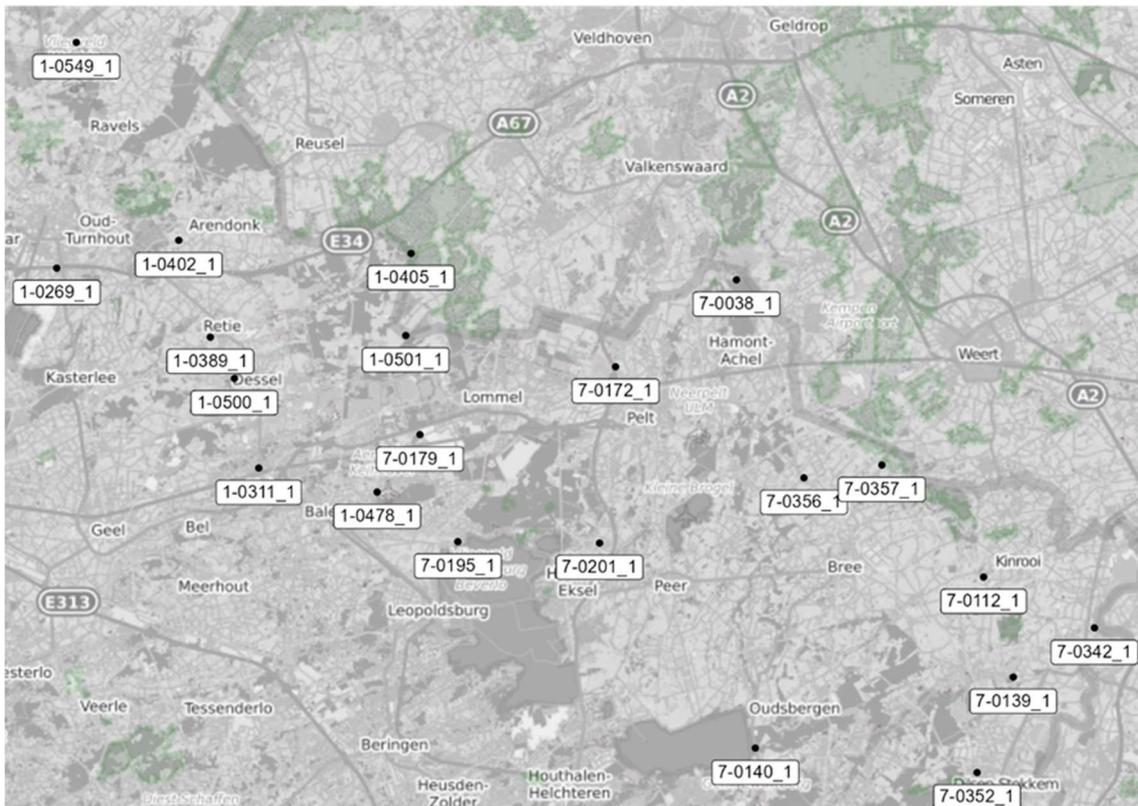


Figure 3.7 Map with the locations of the 21 VMM indicator wells. Background data: © OpenStreetMap contributors.

The groundwater recharge calculations used here were performed in the framework of the work by Lu (in prep) and the Future floodplains project (<https://www.futurefloodplains.be/>). The approach builds further on that of Di Ciacca (2020), and develops a database of recharge time series, based on a set of process models representing all possible relevant combinations of land cover, soil type, and groundwater table depth. Different from the work of Di Ciacca is the inclusion of the curve number method (Cronshey 1986) for accounting for surface runoff. The numerical code used here is HYDRUS-1D (Simunek et al. 2005) which solves the modified form of the Richards' equation. Groundwater recharge was simulated for the period 1990-01-01 to 2019-12-31, with daily time steps, for the main land cover and soil type combinations in the Zwarte Beek, Mombeek and Dijle catchments, while here we use the resulting database for the Zwarte Beek catchment, as it is located closest to the H3O-plus project area.

Figure 3.8 provides an idea on the workflow, which next to the HYDRUS-1D model also involves reference evapotranspiration calculations (FAO 56 Penman-Monteith, using the REF-ET software (Allen 1992)) from meteorological time series like air temperature, relative humidity, wind speed etc., and a canopy water balance model for different vegetation covers (Leterme et al. 2012). In the HYDRUS-1D model, root water uptake

parameterization of these vegetation covers is also implemented, as well as that of the unsaturated soil hydraulic properties, which is based on the Aardewerk database (Van Orshoven et al., 1988). The top boundary of the soil column in HYDRUS-1D is exposed to the transient forcing by throughfall, potential evaporation and potential crop transpiration. The bottom boundary is implemented as a constant head boundary (free drainage) with a constant groundwater depth, up to 5 m below the soil surface.

A more detailed description is considered out of scope here. The general idea is just that this approach gives us a time series database, of recharge in function of land cover, soil type and water table depth. As it is a process modelling approach, the recharge derived here results from all kinds of non-linear processes in the unsaturated zone. Important to realize, however, is that these are all forward model simulations, working with best estimates of the parameter values, without any calibration whatsoever. Hence, it is likely that some rescaling of the resulting time series is for instance required for a specific site.

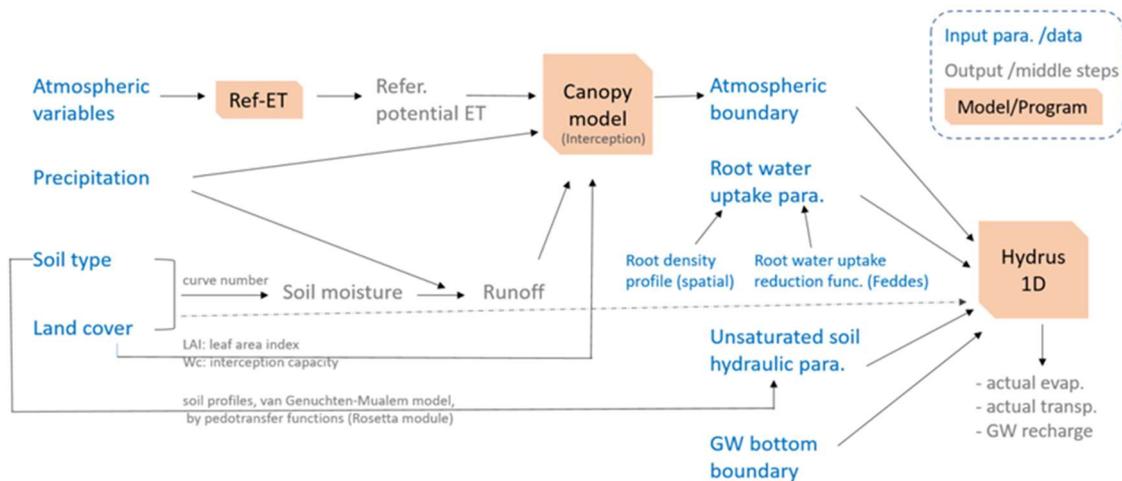


Figure 3.8 Workflow for the recharge process modelling.

For identifying the different impulse-response function model architectures, we use a four letter abbreviation. The first letter represents the model inputs, where R is the meadow on sand with a groundwater table at 2 m depth recharge time series from the process model results database, P is the combination of the first three principal components of that database, and M means the more traditional meteorological time series of precipitation and evapotranspiration. Non-linear recharge estimates, other than those provided by the process model results database (so similar to the approaches by Long (2015) and Collenteur et al. (2021)), were initially also foreseen, but given time constraints, they are not included here. Hence, the second letter in this document is always L, for linear estimation of the impulse. The third letter indicates whether we consider constant, time-invariant impulse-response function (C), or a time-variant impulse-response function (T), while the fourth and final letter is N in case no attempt at addressing temporal correlation in the residuals is made, while G signifies that the model error is simulated explicitly using a Gaussian process.



In all of the investigated architectures, the convolution output is considered to represent the explained part (by recharge or meteorology) of the groundwater level time series. What is considered to be the unexplained part depends on the approach:

1. In case nothing is foreseen to address temporal dependence of residuals, the difference between the convolution output and the observed levels is considered to represent the unexplained part. This is in fact lumping the model and measurement errors together. For quantifying this unexplained part on the complete time series, linear interpolation is used.
2. In case of the Gaussian process approach, the Gaussian process projection to the full time series represents the unexplained part. This represents the model error and does not include the measurement error.

All possible combinations are tested for exploratory purposes:

- RLCN: single recharge time series with response that is constant in time;
- RLCG: as RLCN but with a Gaussian process;
- RLTN: singel
- RLTG: as RLTN but with a Gaussian process;
- PLCN
- PLCG: as PLCN but with a Gaussian process;
- PLTN
- PLTG: as PLTN but with a Gaussian process;
- MLCN
- MLCG: as MLCN but with a Gaussian process;
- MLTN
- MLTG: as MLTN but with a Gaussian process.

A brief overview of the different resulting architectures is provided in Appendix C. This provides a clear idea on the different approaches and how they compare to each other.

Time series analysis is performed for both the explained and unexplained parts from the different model architecture results, as we are interested in both explained (by meteorology) as well as unexplained (model error and anthropogenic effects) trends. The procedure implemented here consists of the following steps:

1. First, STL decomposition is performed using an implementation robust to missing values (Hafen 2016), with seasonal and trend windows of seven years, and quadratic polynomials for the local regression. For the unexplained part, the results are omitted for dates without any level observation within a month of the considered date, under the assumption that it is impossible to know what type of anthropogenic impact may have been in effect for such gaps in the observation time series. This operation results in trend, seasonal and remainder components for both the explained and unexplained parts of the time series.
2. Next, the trend and seasonal strength STL features of the explained and unexplained parts are calculated, according to their definition in Hyndman & Athanasopoulos (2021), mainly for visualization purposes (see Section 2.2).



3. Finally, breakpoint detection is performed on a smoothed version of the derivative of both explained and unexplained trend components, using the method by Zeileis et al. (2002, 2003) for dating structural changes in linear regression models, again masking some data around gaps in the time series.

The R packages specific to the used methodology have been explicitly mentioned. Other packages used in the workflow and pre- and post-processing of results deserve to be mentioned as well: *tidyverse* (Wickham et al. 2019), *here* (Müller, 2020), *fs* (Hester & Wickham 2020), *arrow* (Richardson et al. 2021), *sf* (Pebesma 2018), *sfheaders* (Cooley 2020), *nomnoml* (de Vries & Luraschi 2020), *RcolorBrewer* (Neuwirth 2014), *future* (Bengtsson 2020), *patchwork* (Pedersen 2020), *ragg* (Pedersen & Shemanarev 2021), and *ggridges* (Wilke 2021) packages.

### 3.1.2 Germany

The German part of the H3O-Plus area lies in the land North Rhine Westphalia.

#### 3.1.2.1 Organisation

The department for Nature, Environment, and consumer protection of North Rhine-Westphalia (LANUV, Landesamt für Natur, Umwelt und Verbraucherschutz Nordrhein-Westfalen, <https://www.lanuv.nrw.de/>) is an important organisation in the groundwater quantity monitoring of the land of North Rhine-Westphalia (NRW). In the Roer Valley Graben, additional monitoring is carried out by the waterboard Erftverband (<http://www.erftverband.eu/>), the mining industry, and public water supply organisations. Monitoring data is made available through the webportal ELWAS (<http://www.elwasweb.nrw.de/>).

There are 6683 monitoring wells in the Roer Valley Graben of which 3698 are active.

#### 3.1.2.2 Analysis method

Various methods for trend assessment are used:

- Regression analysis;
- Correlation analysis;
- Statistical testing;
- Multi-channel Wiener-filter analysis.

In the regression analysis a regression line is fitted through the groundwater head observations and the significance of the resulting trend is determined (Figure 3.9).

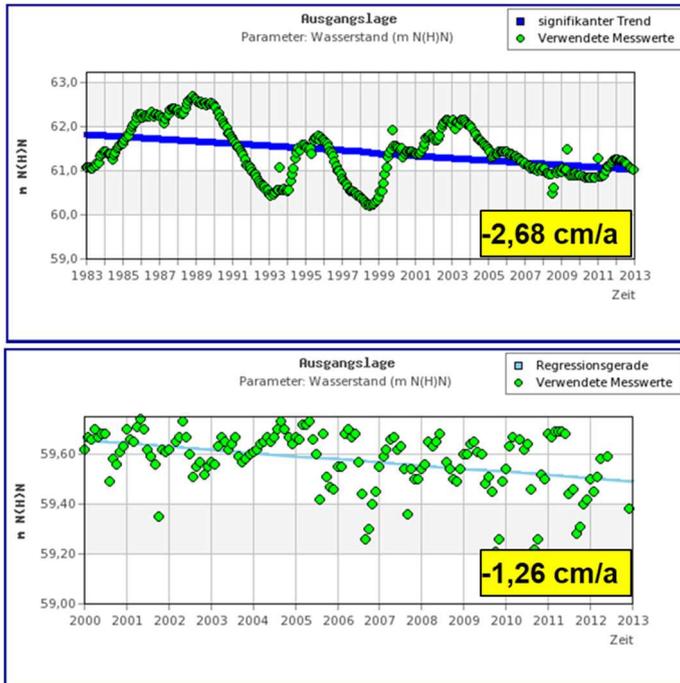


Figure 3.9: Two examples of regression analysis with (top) and without (bottom) a statistically significant trend).

The correlation analysis does not examine the existence of a trend in individual time series but shows whether multiple time series have the same behaviour (Figure 3.10).

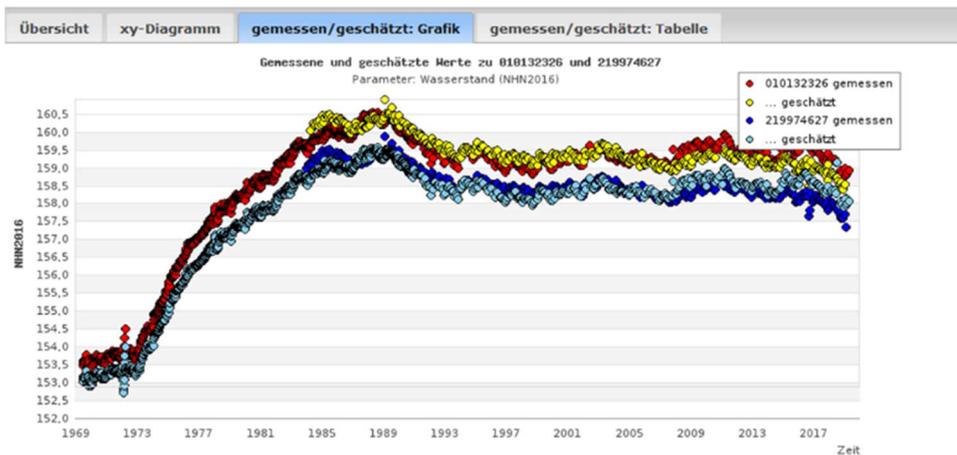


Figure 3.10: Example of correlation analysis.

The statistical test examines the relation between a target time series and reference time series in more detail (Figure 3.11).

<b>Zielmessstelle:</b>	58GP018104	010446904	
<b>Referenzmessstellen:</b>	BRACHT 017 DANSWEILER DUELKEN 0059	080100776 279373211 289001316	<span style="color: green;">—</span> <span style="color: yellow;">—</span> <span style="color: purple;">—</span>
<b>Stationäre Phase:</b>			
von	15.01.1990		
bis	08.12.1999		
Datum:	06.02.2014		

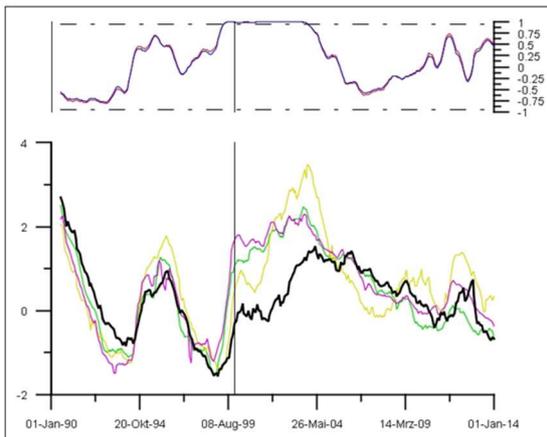


Figure 3.11: Example of statistical test.

The Multi-channel Wiener-filter also uses reference time series to get more insight in a target time series (Figure 3.12).

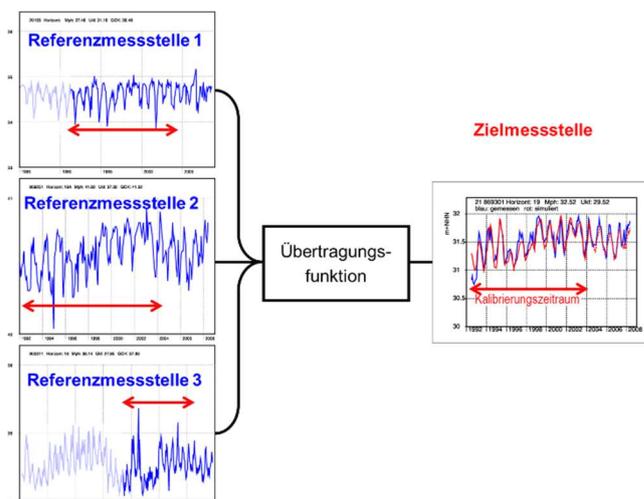


Figure 3.12: Example of multi-channel Wiener-filter analysis.

This enables a more refined determination of the trend compared to the regression analysis (Figure 3.13).

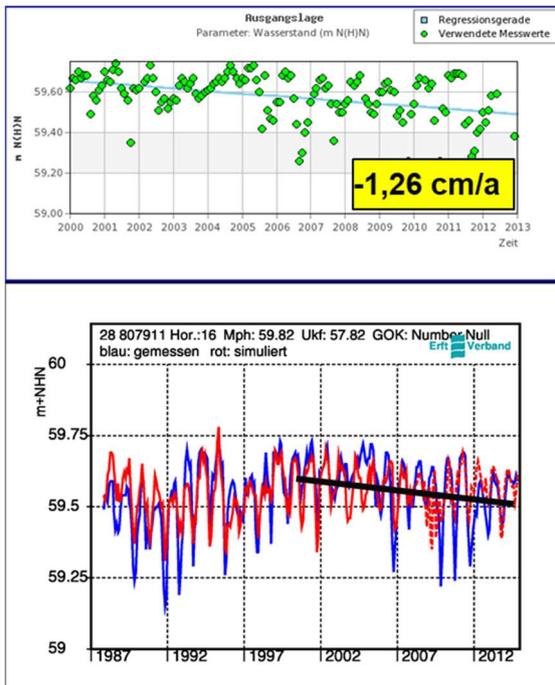


Figure 3.13: Example of trend from regression analysis and Wiener-filter analysis.

The regression analysis and correlation analysis are simple to apply but the conclusions are not be used easily. The statistical test and the Multi-channel Wiener-filter analysis are more complex to apply, but lead to more sophisticated conclusions, that are more suitable for the groundwater management.

### 3.1.2.3 Visualization method

The web interface of the online groundwater database of the land North Rhine-Westphalia (<https://www.elwasweb.nrw.de/elwas-web/index.jsf>) has the option to visualize groundwater head time series (Figure 3.14).

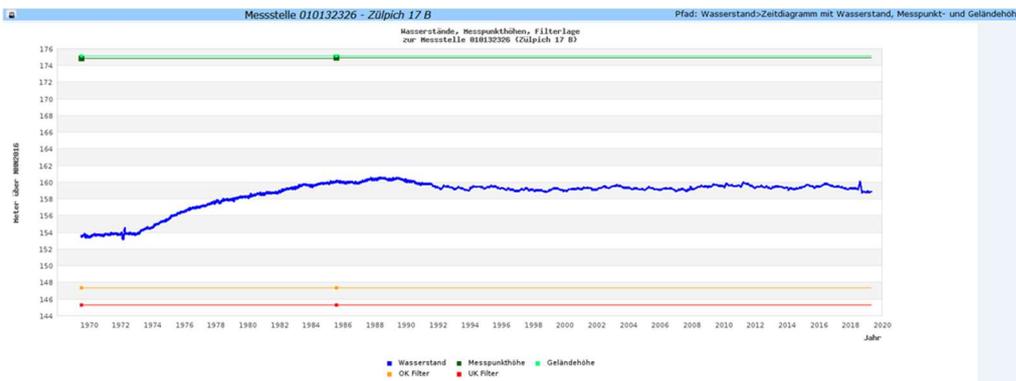


Figure 3.14: Example of time series in the ELWAS web application.

Figure 3.15 shows a visualization of the status of groundwater heads with respect to targets for realization of ecological conditions.

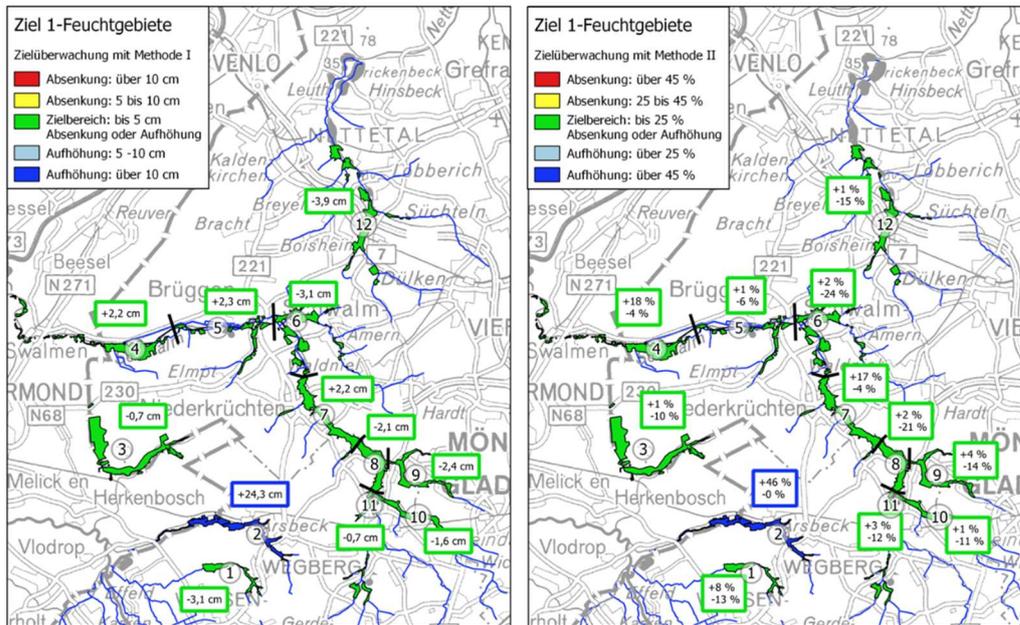


Figure 3.15: Example visualisation of realization of target heads for the Garzweiler monitoring in 2018 for the statistical test (left) and the multi-channel Wiener-filter analysis (right).

### 3.1.3 Netherlands

The Dutch part of the H30-Plus area is located in the provinces of Noord-Brabant and Limburg.

#### 3.1.3.1 Organisation

The Provinces are responsible for the general purpose of monitoring of groundwater heads. The waterboards and municipalities have monitoring networks to support their own water management tasks. They usually put more emphasis on the phreatic groundwater table. Nature organisations have dedicated groundwater monitoring especially for Natura2000 and Habitat directive areas. Water companies also monitor groundwater, prescribed by their groundwater abstraction permits and prompted the information need of their own operations. Monitoring on a project basis is carried out for projects with a potential important impact on the groundwater, e.g. large river works and construction sites.

As of the first of January 2021, groundwater heads are part of the National Key Registry of the Subsurface (see <https://basisregistratieondergrond.nl/english>). This means that all government organisations have to submit the head measurements they acquire and they have to consult the registry when information on groundwater heads is relevant. The groundwater data in the registry is publicly available together with the existing data in the national subsurface database (<https://www.dinoloket.nl/en/subsurface-data>).

### 3.1.3.2 Analysis method

In water management and policy development and evaluation in the Netherlands, trends in groundwater levels are separated in natural and anthropogenic changes. Precipitation and reference evaporation (Makkink) are considered natural influences. Anthropogenic influences include land-use changes (e.g. conversion of agricultural to urban land-use), water management changes (e.g. implementation of drainage), groundwater extraction (e.g. for drinking water), and changes of land-use (e.g. higher crop yields in agriculture). However, concern is growing about (ground)water availability since the dry summer of 2018 (in the East and South of the country, 2019 and 2020 were relatively dry also). Therefore, trends of the groundwater levels are also considered independently of the cause.

In policy evaluation and reporting for the EU Water Framework Directive, the natural fluctuation of groundwater levels usually is determined by transfer-noise modelling of groundwater head measurement time series with precipitation and Makkink evaporation as explaining variables. The remaining variation usually is assumed to be caused by anthropogenic influences. The analysis is carried out per groundwater body (Figure 3.16).

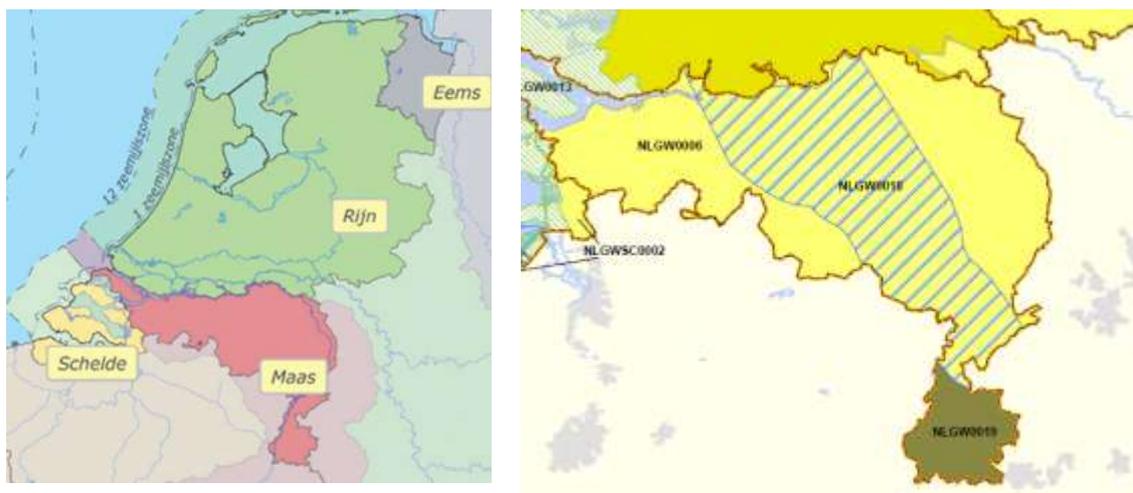


Figure 3.16 River basins in the Netherlands (left) and the groundwater bodies in the Dutch part of the Meuse river basin (right).

Transfer-noise modelling splits the groundwater level time series into a deterministic part and a stochastic part (Figure 3.17). The deterministic part represents the variation due to the specified explanatory variables.

TNO Geological Survey of the Netherlands uses the software Metran (Berendrecht & van Geer, 2016) for the time series modelling (Zaadnoordijk et al., 2019). For the models on the 'groundwatertools' website (<http://www.groundwaterstandeninbeeld.nl>), these are precipitation and potential evapotranspiration. It is possible to include additional influences, like surface water levels or a general trend. The difference between the deterministic part and the measurements is called the model residual.

A noise model is used for the stochastic part. The purpose is to remove the autocorrelation in the residuals. The smaller the time steps between the measurements, the larger the autocorrelation. The existence of autocorrelation decreases the reliability of the model. We use a noise model with an exponential decay. The inverse of the noise model is applied to the residuals to obtain so-called “innovations”.

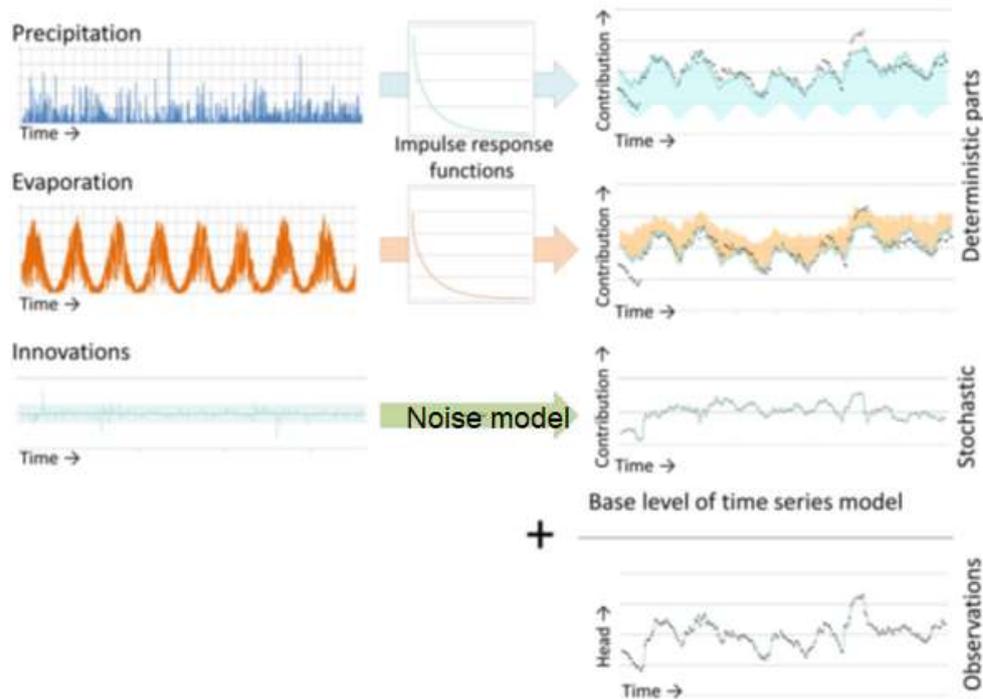


Figure 3.17 Setup of transfer-noise modelling of groundwater head time series in Metran.

The explanatory variables are convoluted with an impulse response function (see e.g. Kreyszig, 2012): the value of each day is multiplied by the response function and the results are summed. An incomplete gamma distribution is used for the impulse response function (Berendrecht & Van Geer, 2016). It has three parameters, a multiplication factor  $A^*$  and two shape parameters  $a$  and  $n$  (Besbes & de Marsily, 1984). For the groundwatertools website, the same function is used for precipitation and potential evapotranspiration except for a factor. This leads to five parameters to be optimized: three of the precipitation response, one evaporation factor, and one noise model parameter. The parameters are determined by a minimization procedure for the innovations.

The resulting time series models are evaluated using model evaluation criteria among which the explained fraction of the groundwater variation (Zaadnoordijk et al., 2019). Three classes are distinguished: bad models, reasonable models, and good models. The bad models are not shown on the website. The analysis in this report uses only the good models.

Additional explaining variables can be used in the transfer-noise modelling of head time series, e.g. when a specific effect needs to be determined. Figure 3.18 shows an example with the water level in the Meuse river as third explaining variable.

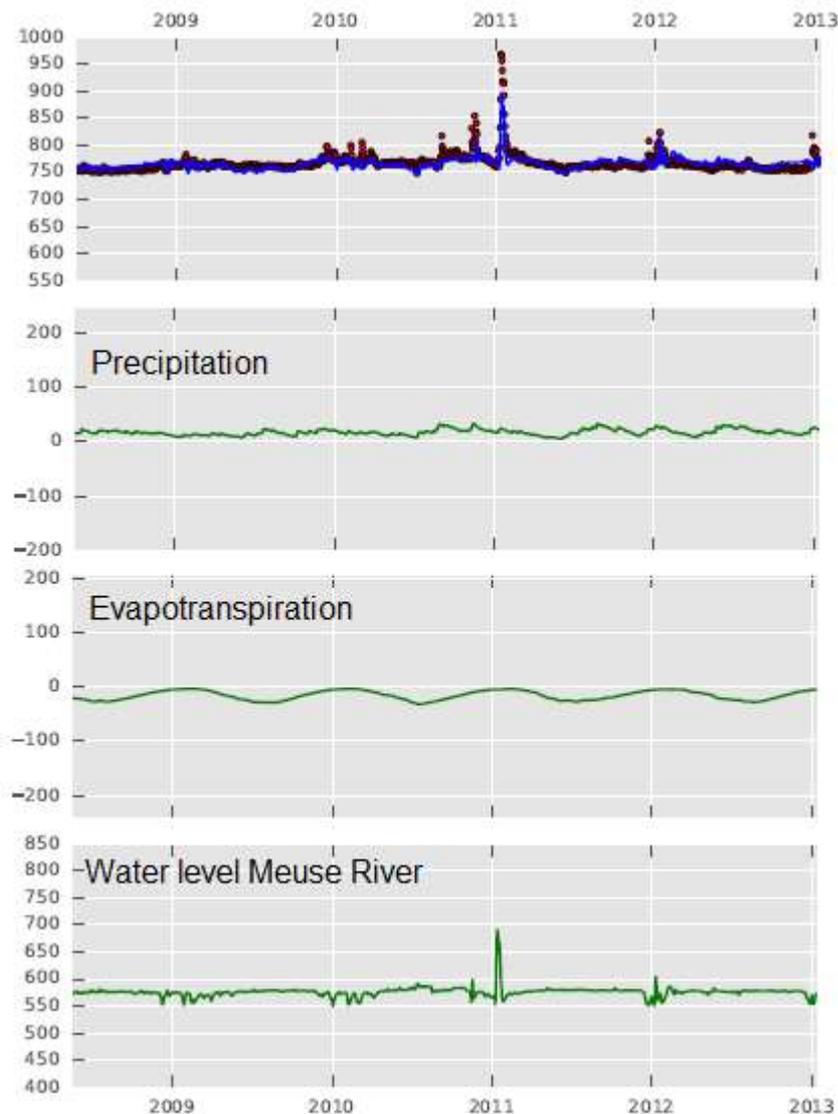


Figure 3.18 Metran model with the river water level of the Meuse as explaining variable next to precipitation and evaporation

### 3.1.3.3 Visualization method

The reports for the Water Framework Directive visualize groundwater heads using 6 year averages of meteo corrected time series. These can be either absolute values, or values relative to a reference period.

Metran can be used for the meteo correction, if Metran can determine a good time series model for the head time series with precipitation and evaporation as explaining variables.

Using this model, absolute values of meteo corrected heads are based on a reference precipitation and evaporation. This so-called ‘structural level’  $\phi_s$  is calculated as:

$$\phi_s = M_{0,P}P_{ref} + fM_{0,P}E_{ref} + b$$

Where  $P_{ref}$  is the reference precipitation,  $E_{ref}$  is the reference evapotranspiration,  $M_{0,P}$  is the total precipitation response of the Metran model,  $f$  is the evaporation factor of the Metran model, and  $b$  is the base level of the Metran model.

The relative values can be determined from the residuals of the Metran model which reflect the head variation which is not caused by precipitation and evaporation. The averages of the residuals are then calculated for six year periods, and are shifted vertically such that the average for the reference period is zero (e.g. Figure 3.19).

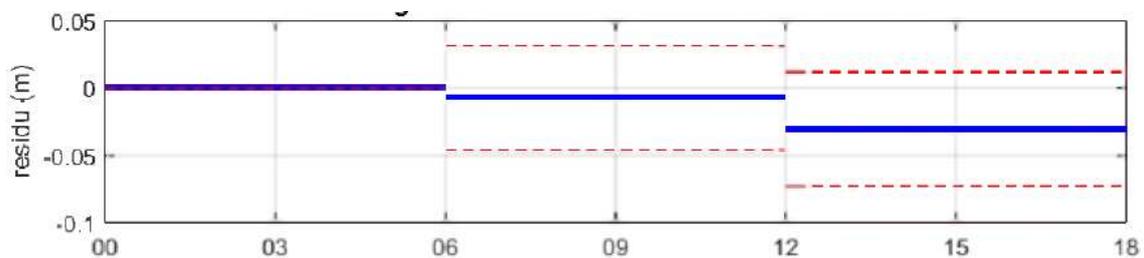


Figure 3.19 Six year averages of the residu of a time series model for B57H0047003 (from: Leunk & van Doorn, 2017).

The public database (<https://www.dinoloket.nl/ondergrondgegevens>) provides a very simple visualization of the time series (Figure 3.20).

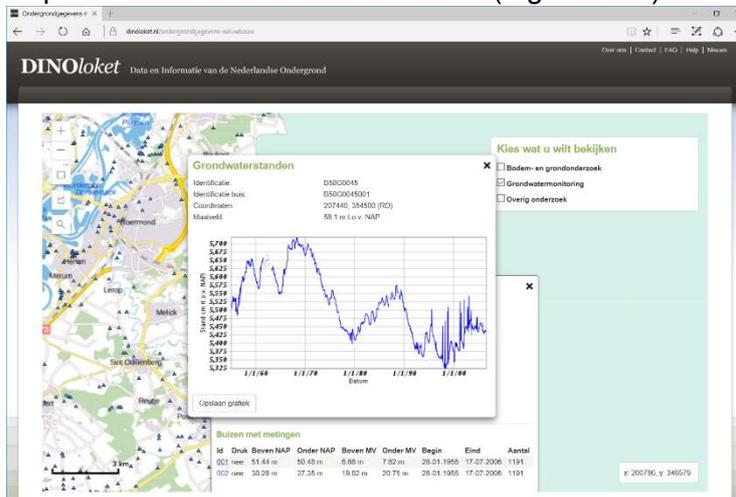


Figure 3.20 Visualization of time series in the National subsurface database.

TNO-GSN makes visualizations publicly available based on Metran models (see <http://www.grondwaterstandeninbeeld.nl>). The visualizations include the contribution of precipitation and evaporation, the regime curve, and various statistics (Figure 3.21 - Figure 3.23) (also see Zaadnoordijk et al., 2019).

## Putlocatie B57H0232

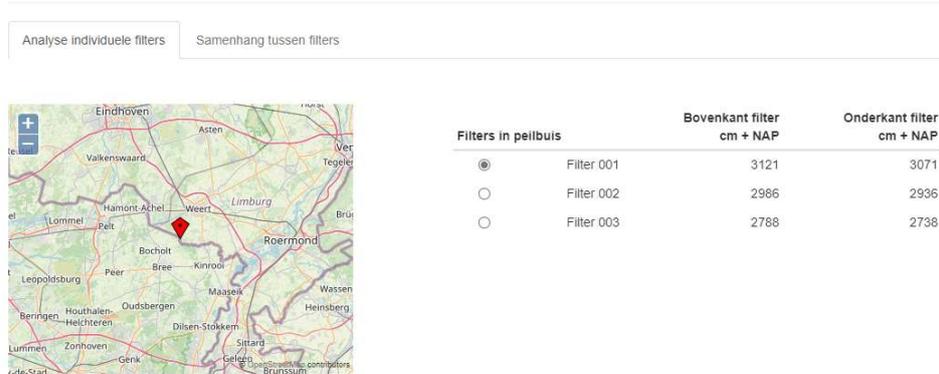


Figure 3.21 overview from tool: location and elevation of piezometers.

The location is shown on a map and the piezometers are shown with their elevation and an option to select another piezometer of the monitoring well. For the selected piezometer, the time series is shown with the option to include the surface level (Figure 3.22).

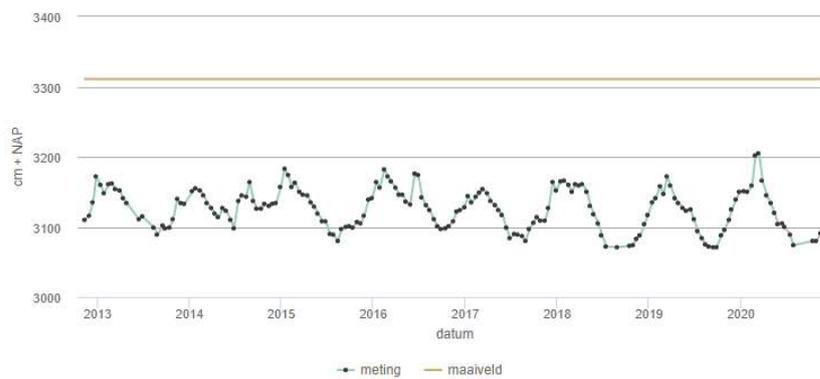


Figure 3.22 tool: time series visualization head with surface elevation.

The measurements of all piezometers can also be viewed in a single graph (Figure 3.23).

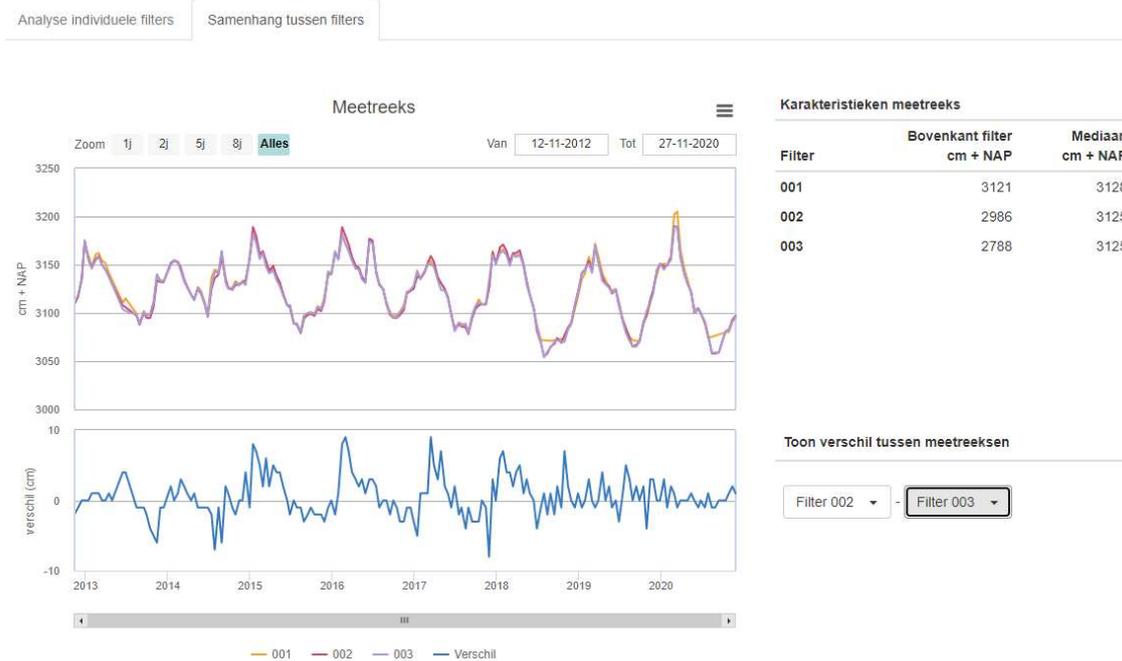


Figure 3.23 visualization in tool of head series from all piezometers of a monitoring well.

The Metran time series model for the head time series together with precipitation and evaporation is evaluated using several quality criteria (Zaadnoordijk et al., 2019). If the model fulfils the basic criteria, the separation of the groundwater variation into contributions of precipitation, evaporation and other influences is shown (Figure 3.24). Also the precipitation and evaporation contributions over time of can be visualized.



Figure 3.24 tool's visualization of influences in time and relative to the variation not explained by Metran.

The impulse response functions can be regarded as signatures of the groundwater system. They are shown as a graph and with some characteristic values such as the total response  $M_0$  and the time to the median response  $t_{50}$  (Figure 3.25).

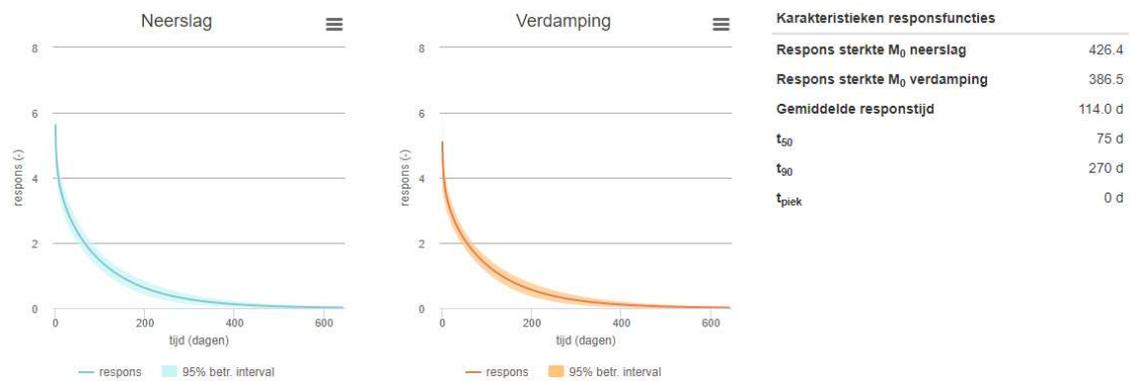


Figure 3.25 visualization of impulse response functions in tool.

The advanced information that can be shown are related to the model parameters and the statistics that are used for the quality criteria to judge the model (Figure 3.26).

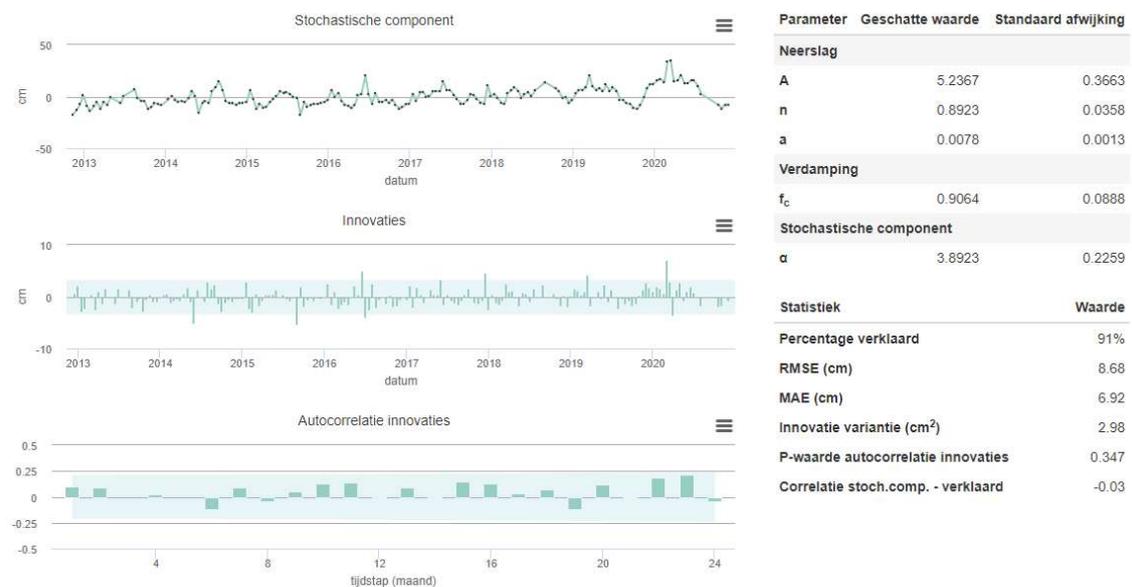


Figure 3.26 visualization of residues, innovation and autocorrelation innovation for check of Metran model.

If the model fulfils all quality criteria, then also a regime curve is shown (Figure 3.27). The regime curve gives the long term median variation of the head over the year together with percentiles.

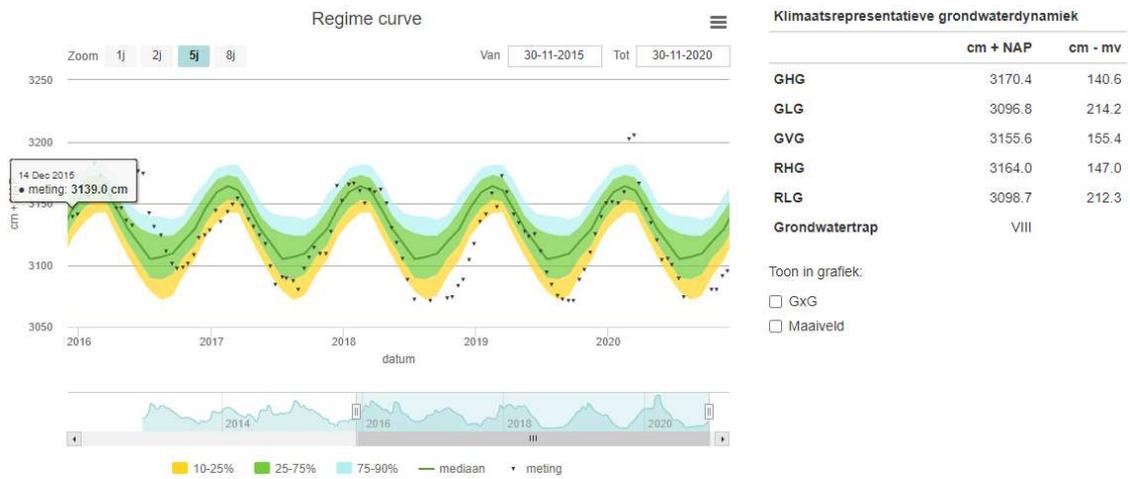


Figure 3.27 tool visualization: regime curve.

The regime curve allows to rate measurements with respect to the normal seasonal variation. The percentiles together with the absolute value of the measurements provide extra insight in the occurrence of extreme groundwater levels.

## 3.2 Heads for entire H3O-PLUS area

For the heads of the entire H3O-PUS area, Metran has been used together with 3-d visualization of characteristics of time series models. More detail is given in the paragraphs below.

### 3.2.1 *Metran: transfer noise modelling of groundwater head time series*

Metran (see Figure 3.17 and Subsubsection 3.1.3.2) has been used for the modelling of groundwater head time series using the meteorologic information collected in the project database (see Chapter 2).

#### 3.2.1.1 *Metran output used*

Two characteristics of the precipitation response will be used: the total response (the final unit step response, which can be calculated as the zeroth moment of the impulse response function) and the average response time (equal to the ratio of the first moment and the zeroth moment). These characteristics for the precipitation impulse response reflect the groundwater system.

Furthermore, the residuals are investigated for trends. The residuals are the differences between the model and the measurement and are assumed to represent the anthropogenic influences on the groundwater (cf. Figure 3.28).

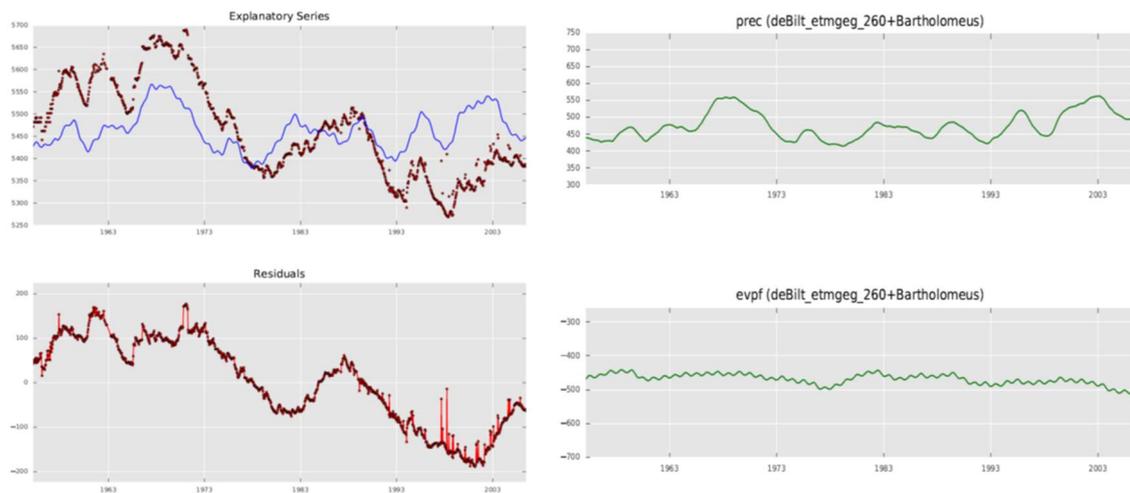


Figure 3.28 trend in residu of Metran model with precipitation and evaporation for piezometer B58G0045002.

However, the presence of a trend in the head series which is not related to the explanatory variables makes it more difficult for Metran to determine a good time series model. The chance of determining a good time series model increases when the trend is added as an explanatory variable. This is illustrated by Figure 3.28 and Figure 3.29. The model of Figure 3.28 with only precipitation and evaporation cannot explain the large trend in the measurements, and thus the residuals (difference between models and

measurements) are quite large. Figure 3.29 shows a model that also has a trend and now the residuals are smaller and lie more nicely around zero.

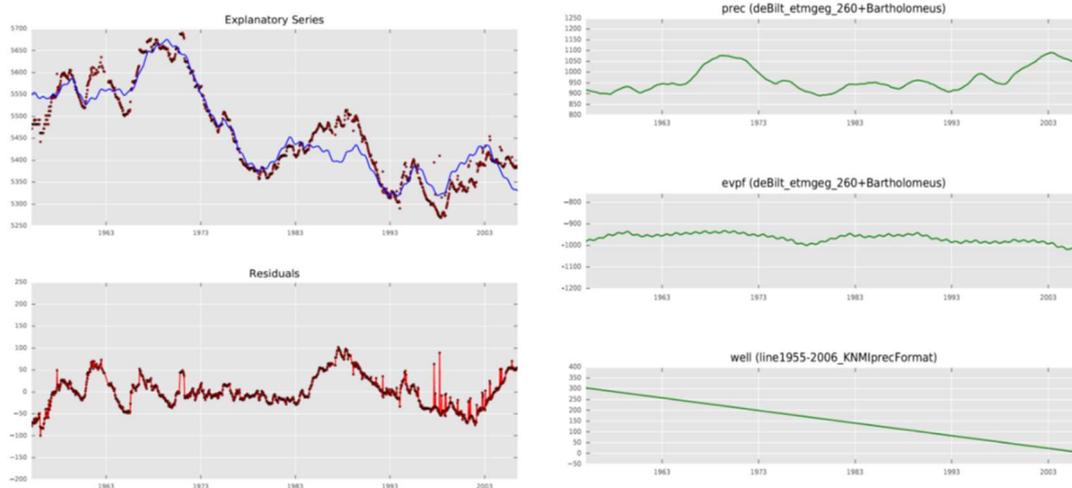


Figure 3.29 Metran model with precipitation, evaporation, and trend for piezometer B58G0045002.

### 3.2.1.2 Temperature instead of evapotranspiration in Metran

A comparison has been made between the use of evapotranspiration and temperature in addition to precipitation in Metran models, because no evapotranspiration data is available for the German part of the H3O-PLUS area.

The time series from the Dutch national database (available through <https://www.dinoloket.nl/en/subsurface-data>) for the period 1994-2019 inside the H3Oplus area have been selected for the comparison. A total number of 3674 time series were available, with an average of 2400 groundwater head values per series (minimum 38, maximum 7094).

These timeseries have been used to compare time series models based on precipitation and Makkink evaporation with models having precipitation and average daily temperature as input.

The models with Makkink evaporation have been run with the option built into Metran to automatically select precipitation and evaporation data from the nearest precipitation or weather station of the Royal Dutch Meteorological Institute (KNMI).

The models with temperature have been run with precipitation also from the nearest KNMI precipitation station, but with the temperature from the KNMI weather station in Eindhoven. The results are shown in Appendix B.

### 3.2.2 Trend analysis of groundwater head time series

In addition to the Metran models, the groundwater head series themselves have been tested for trends. The Mann-Kendal test has been used for the existence of a significant trend. If a significant trend exists, it was quantified using the Sen procedure, which is insensitive to outliers.



Trends have been determined for the period 2005-2020 and the period 1995-2010. A few time series with good Metran models have been selected, for which the trends in the contribution from precipitation and evaporation as well as the trend in the model residue have been determined. The model residue is assumed to be equal to the contribution of other (anthropogenic) influences. These trends have been compared to determine the cause of the trends in the groundwater heads.



## 4 RESULTS

### 4.1 Analysis per country

This chapter presents the results obtained with the national procedures that have been presented in section 3.1.

#### 4.1.1 Belgium

##### 4.1.1.1 Individual model architectures

This section provides an overview of the results obtained with the different model architectures, using an example well, namely 1-0389\_1. The same figures for all other wells and architectures are available in appendix. The figures here contain a lot of information, and hence some explanation is required first at this point: The observed levels (black points), explained and unexplained parts (trend + seasonal + remainder), and their trend and trend + seasonal components are displayed in subfigure (a). Breakpoints are indicated as darker green dots on the trend lines. The light gray line in the panel for the explained part represents the sum of the explained and unexplained parts here. Auto-correlation functions are provided in subfigure (b), where the residual is only relevant in case of the Gaussian process approach (and should preferably exhibit no, or little, temporal auto-correlation). An overview of different variances is provided in subfigure (c), where the explained versus unexplained variances, as well as their different components, can give an idea on the model fit.

Do note that subfigures (b) and (c) contain no information on the residual for the architectures without a Gaussian process, as the unexplained part basically represents the model residual (with linear interpolation in between groundwater level observations).

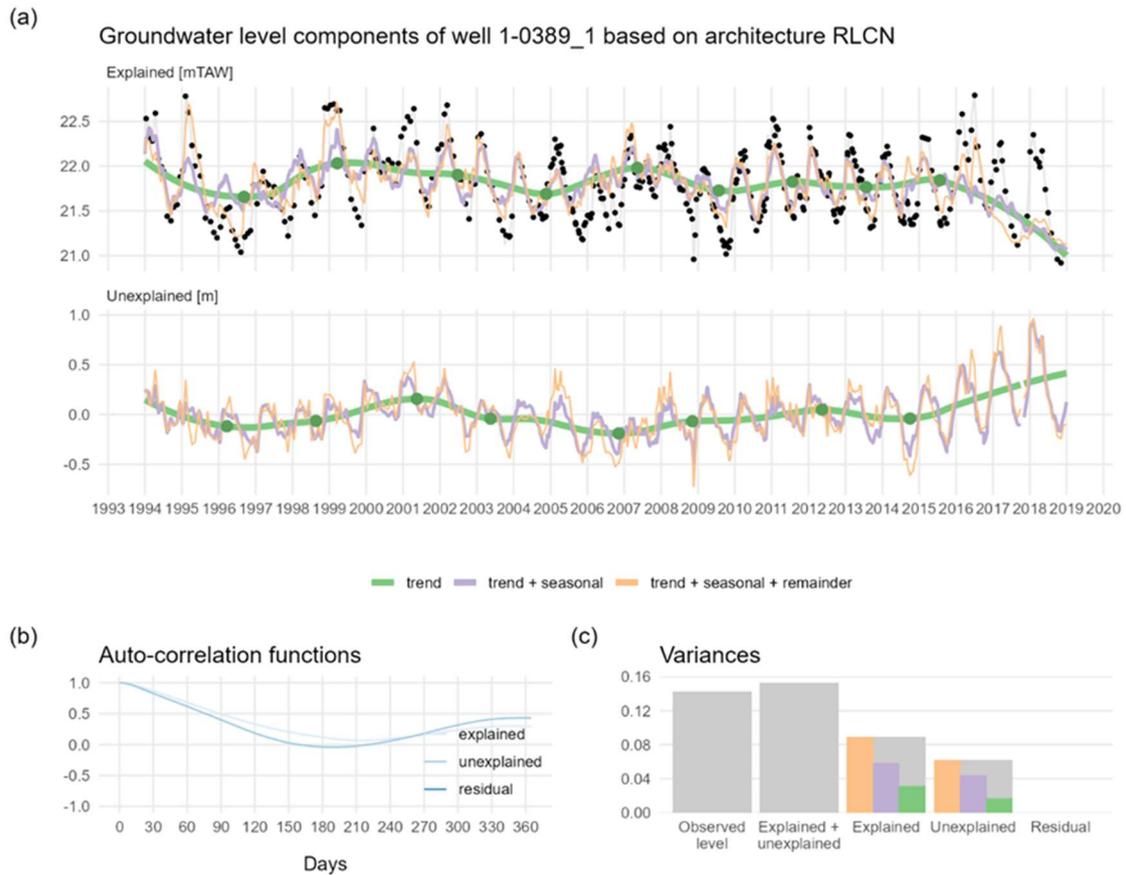


Figure 4.1 Results for the RLCN model architecture, for example well 1-0389\_1.

The RLCN model architecture results are displayed in Figure 4.1. While already a large part of the time series is explained by the recharge for meadow on sand with a groundwater table at 2 m depth, the seasonality of the unexplained part is clearly of the same order of magnitude. This may indicate that the selected recharge time series may not be the most adequate one for this well. The auto-correlation functions clearly show there is autocorrelation in the model error (here the unexplained part), and hence suggest the results may potentially be biased (although this bias is likely limited with regularly spaced observations). The variances suggest that indeed the explained and unexplained variances are not very different, and potentially better fits may be obtained.

The RLCG model architecture results are displayed in Figure 4.2. Here apparently something went wrong, and the Gaussian process takes up most variability in the time series, at the cost of the explained part, which is basically absent. This may indicate stronger regularization is required for the Gaussian process variance (i.e. force it more to lower values) than currently applied. Alternatively, regularization specifically for the seasonal part of the Gaussian process may be appropriate as well (under the assumption that all seasonality should be explained by meteorology, which in fact may not be true), but that would require the time series decomposition to be integrated in the model. The results do show however, that the residual auto-correlation is basically removed by the Gaussian process, except maybe for the first few lags, but the auto-correlation estimates there are very erratic because of the low number of data points involved in the calculation (i.e. only observation dates are used, while the functions for the explained and unexplained parts are smoother because of the continuous time series).

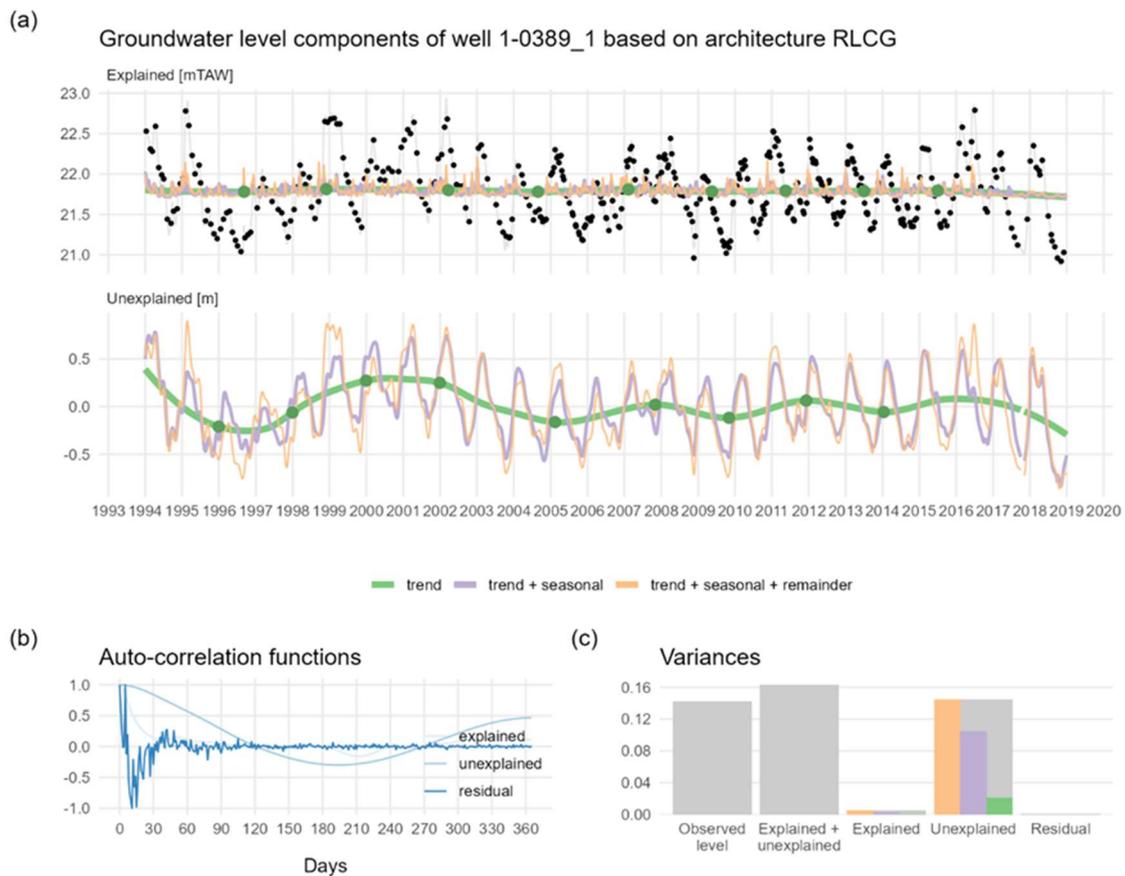


Figure 4.2 Results for the RLCG model architecture, for example well 1-0389\_1.

The RLTN model architecture results are displayed in Figure 4.3. The difference with the RLCN results is however very small, which indicates time-variance of the IRF is not really supported by the data in this case, or cannot compensate for any bias caused by the selection of a single recharge time series for specific conditions.

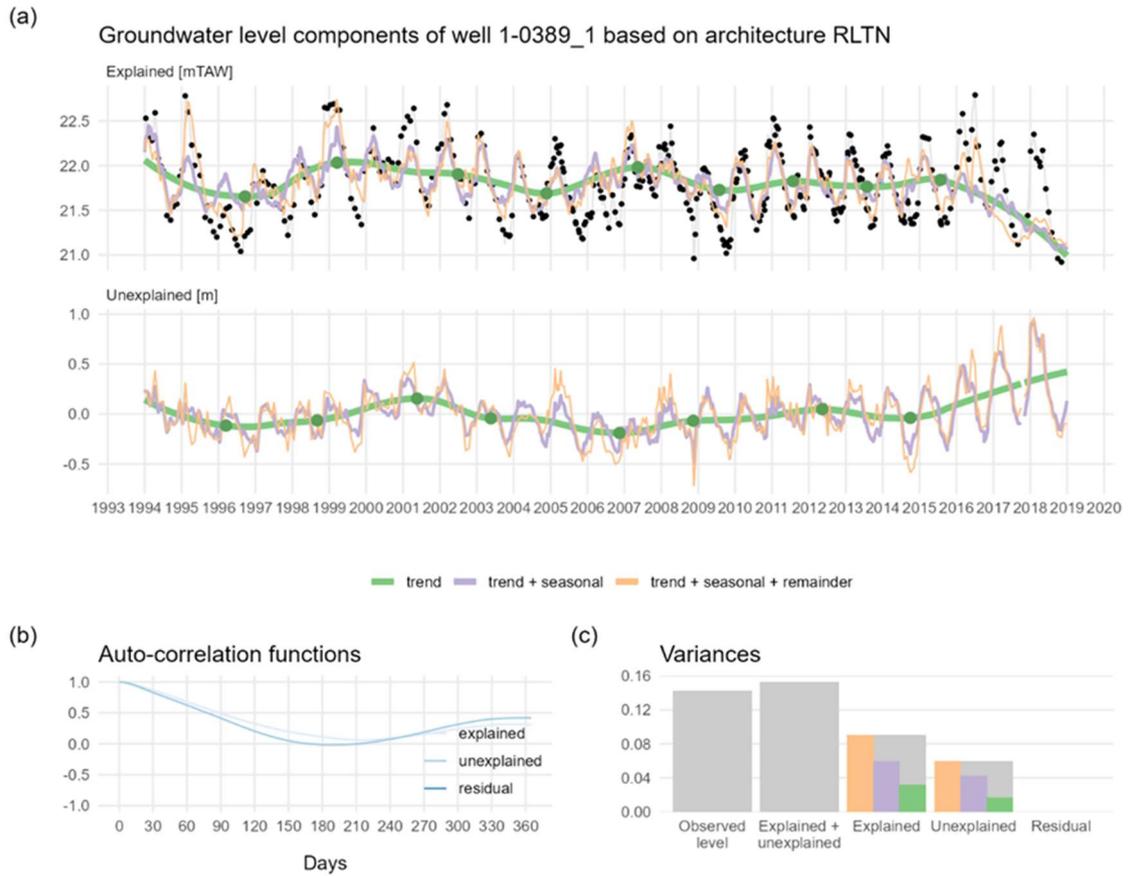


Figure 4.3 Results for the RLTN model architecture, for example well 1-0389\_1.

The RLTG model architecture results are displayed in Figure 4.4. This model does not perform as well as the RLCN and RLTN architectures, as evidenced by the smaller variance of the explained part, than that of the unexplained part. The time-variance apparently prohibits the explained part to complete deteriorate in this case, as happened with the RLTN architecture. Again, some changes to the Gaussian process model approach may be required here, but it does remove the auto-correlation of the residuals.

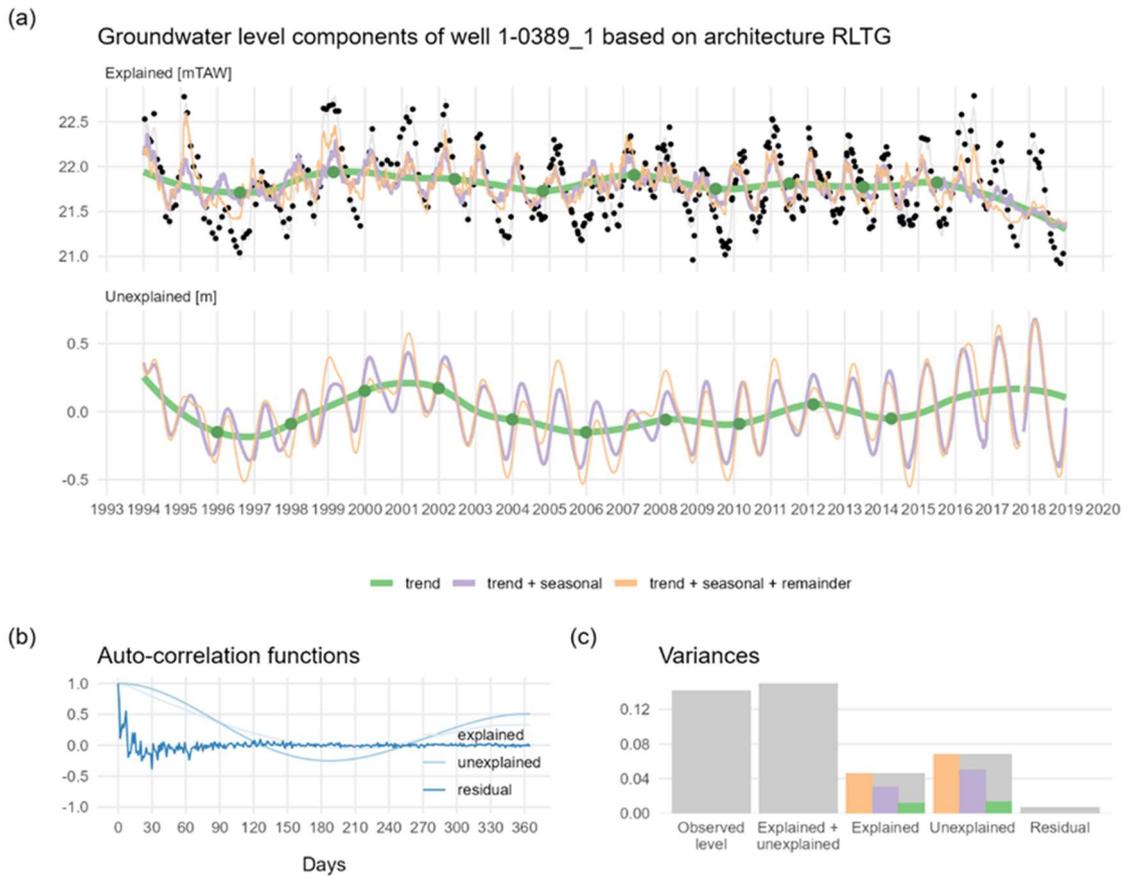


Figure 4.4 Results for the RLTG model architecture, for example well 1-0389\_1.

The PLCN model architecture results are displayed in Figure 4.5. The increased flexibility in determining recharge by use of the principal components of the full dataset apparently results in a considerable increase of the explained variance. Also, the seasonality in the unexplained part is less obvious than in the previous architectures, except maybe in the last few years. Hence, this seems to be a better approach for leveraging a recharge database. Of course, selection of recharge based on specific conditions for a specific well could be done as well, but for the current exercise, land cover, soil type and groundwater table depth for the investigated wells was assumed to be unknown. Moreover, if a larger infiltration area affects the groundwater levels of a certain well, the land cover and soil type at the well location may not even be the most appropriate one, and this approach may still be superior.

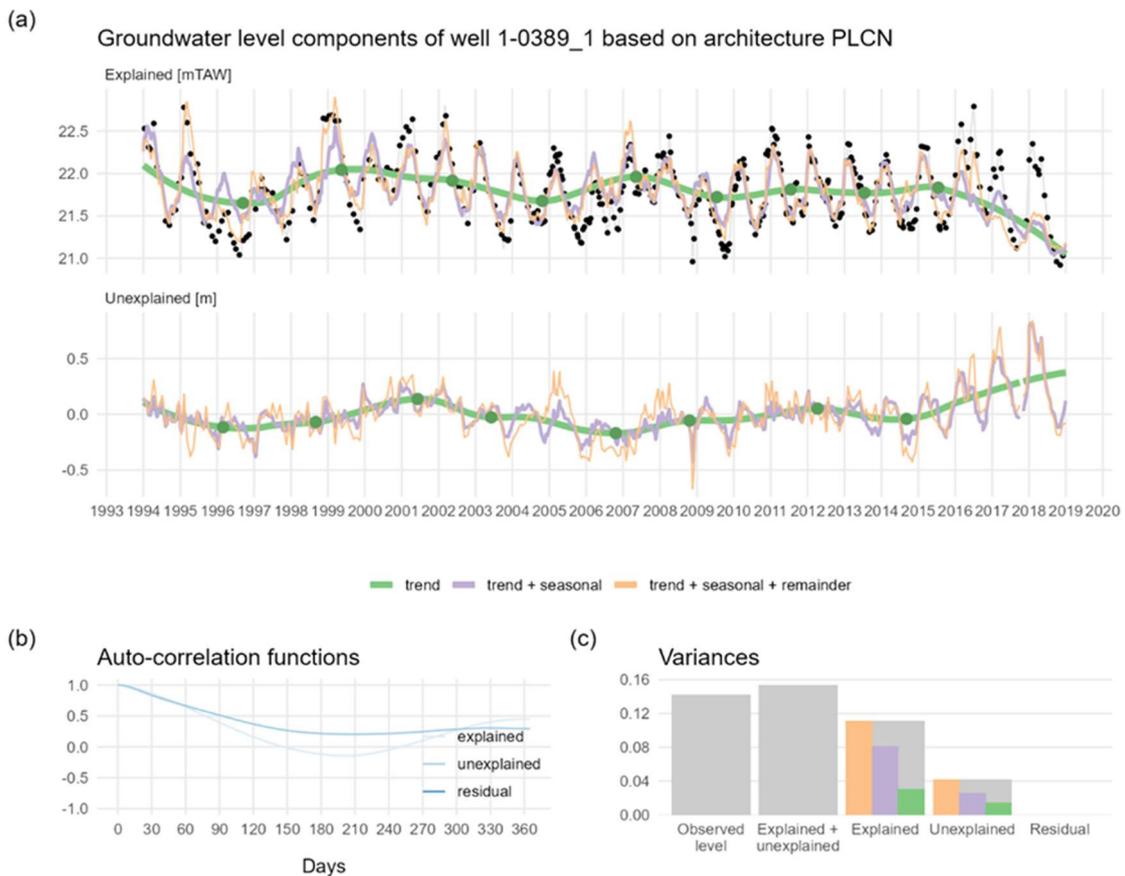


Figure 4.5 Results for the PLCN model architecture, for example well 1-0389\_1.

The PLCG model architecture results are displayed in Figure 4.6. The model performance seems again worse here, likely again due to the Gaussian process that is introduced. The unexplained part again clearly exhibits seasonality, and the variance of the explained part is reduced quite a bit.

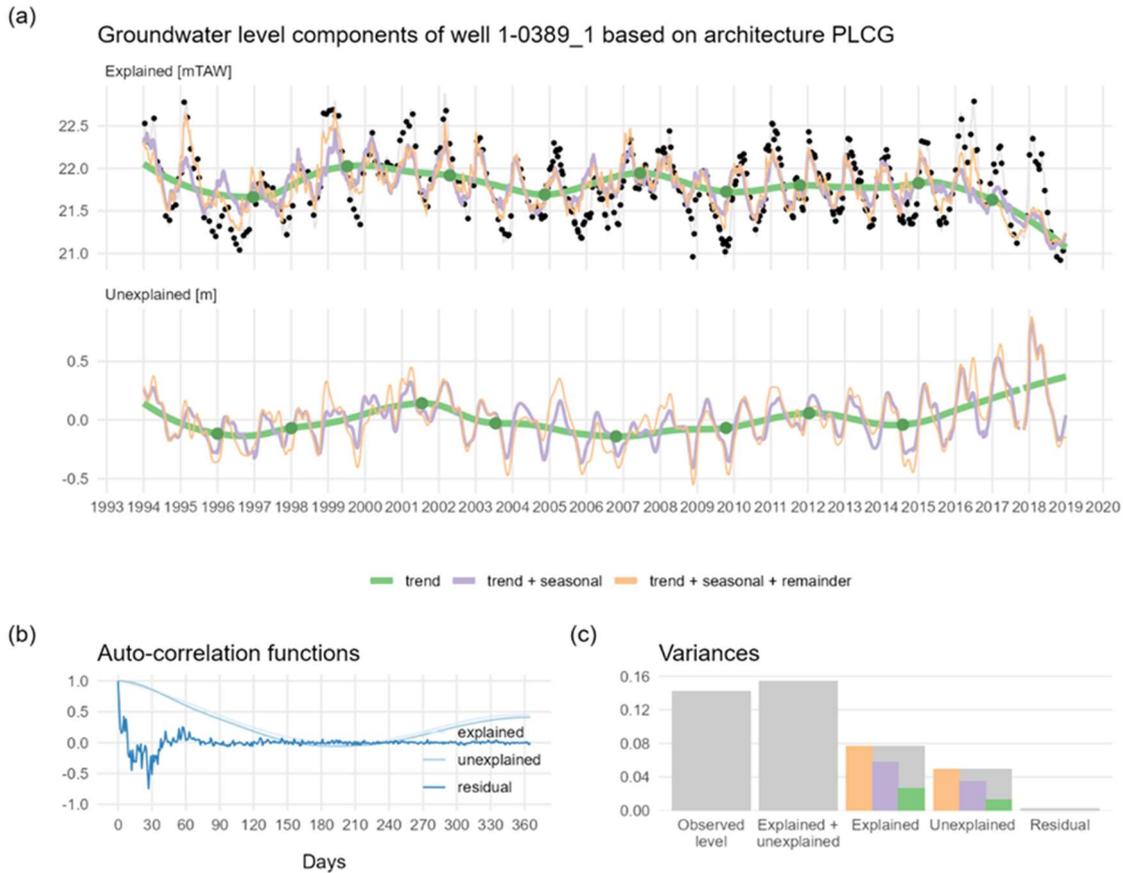


Figure 4.6 Results for the PLCG model architecture, for example well 1-0389\_1.

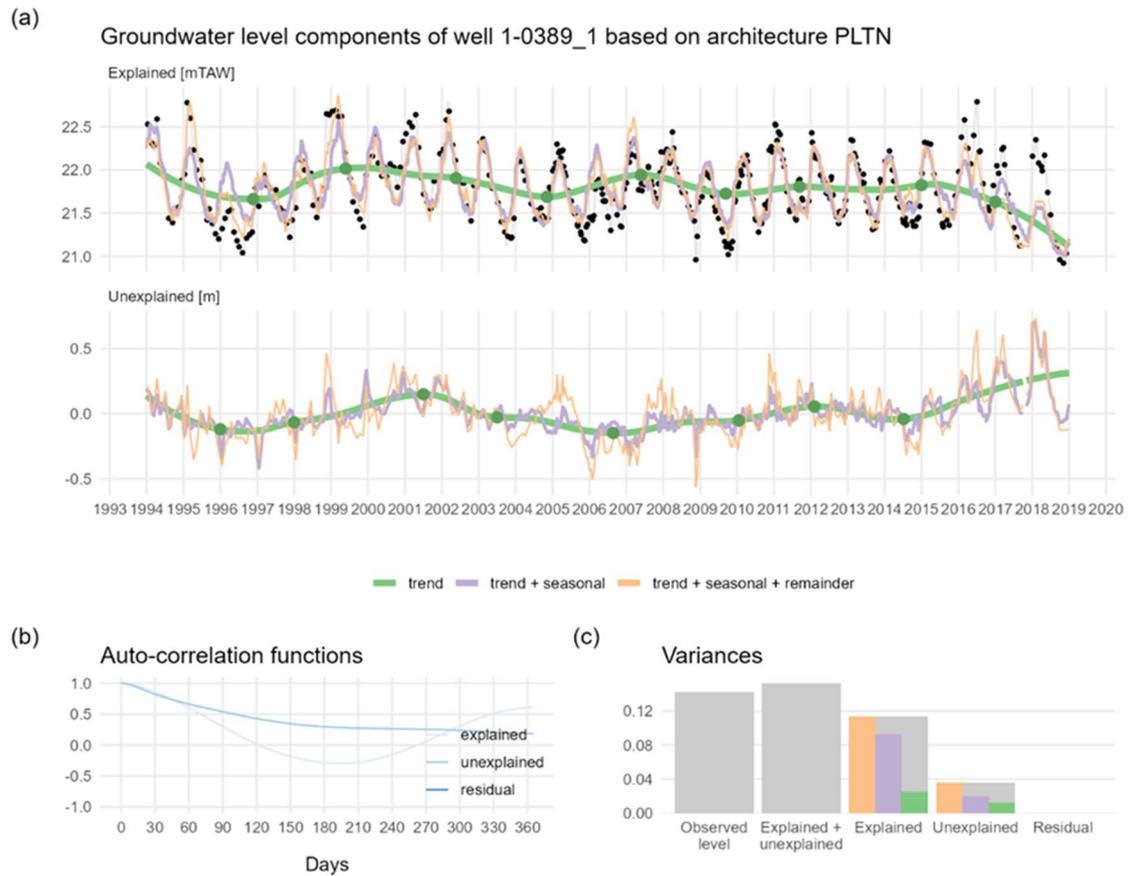


Figure 4.7 Results for the PLTN model architecture, for example well 1-0389\_1.

The PLTN model architecture results are displayed in Figure 4.7. By introducing the time-variance here, the variance of the unexplained part seems to be reduced somewhat, compared to that of the PLCN architecture. The difference is however minor.

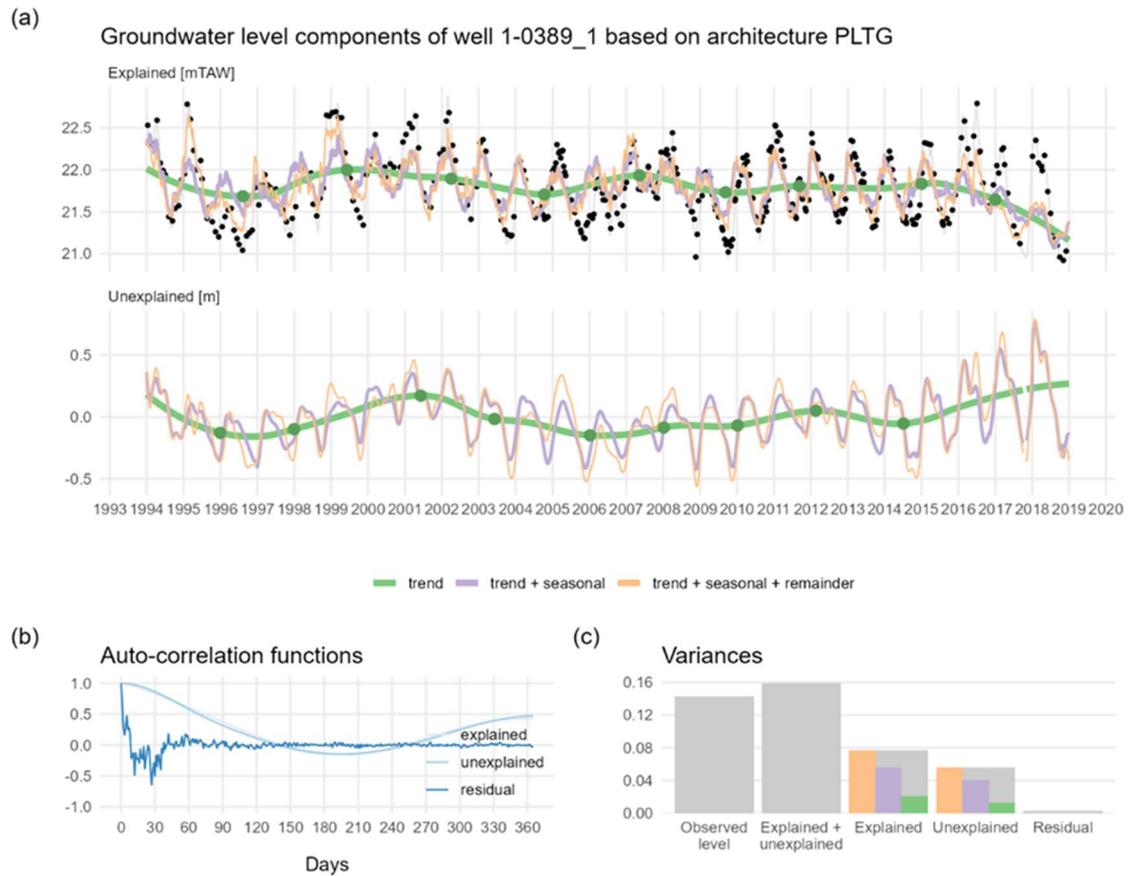


Figure 4.8 Results for the PLTG model architecture, for example well 1-0389\_1.

he PLTG model architecture results are displayed in Figure 4.8. Introducing the Gaussian process model, as well as the time-variance seems to deteriorate the model performance again in this case, with clear seasonality and larger variance for the unexplained part, and lower variance for the unexplained part.

The MLCN model architecture results are displayed in Figure 4.9. The use of the meteorological time series seems to better capture the fluctuations during several years for this particular well. The variance of the explained part is again a bit larger than the maximum obtained for the principal component approach, and seasonality in the unexplained part is not that obvious.

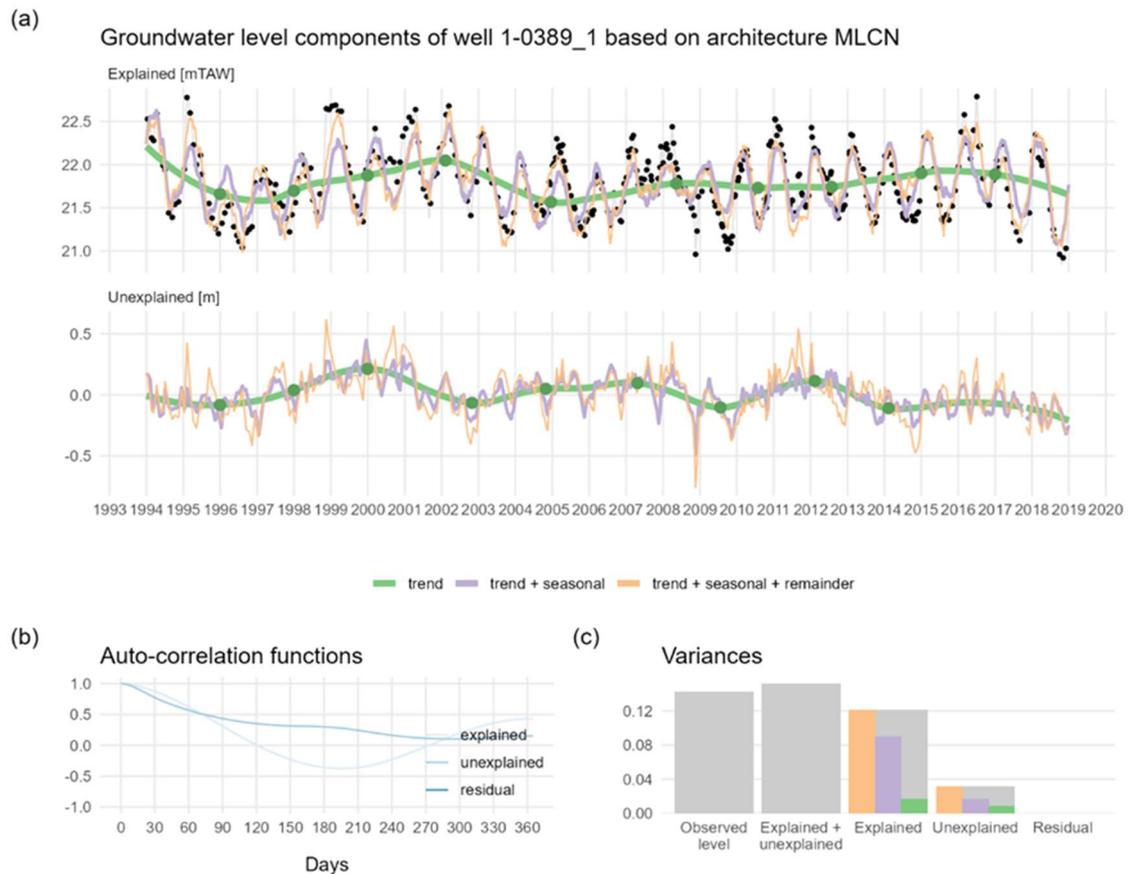


Figure 4.9 Results for the MLCN model architecture, for example well 1-0389\_1.

The striking difference for the last few years, with the previous results, suggests that maybe the meteorological data used for preparing the recharge database, which comes from different stations than the data used here, may not be very representative of what happened in the H3O-plus project area, or at least this well. Moreover, the time series used for the recharge database are composite time series, and not all source data may be equally reliable. Hence, a fair comparison between both approaches is difficult at this stage (which is one of the reasons we keep things exploratory here).

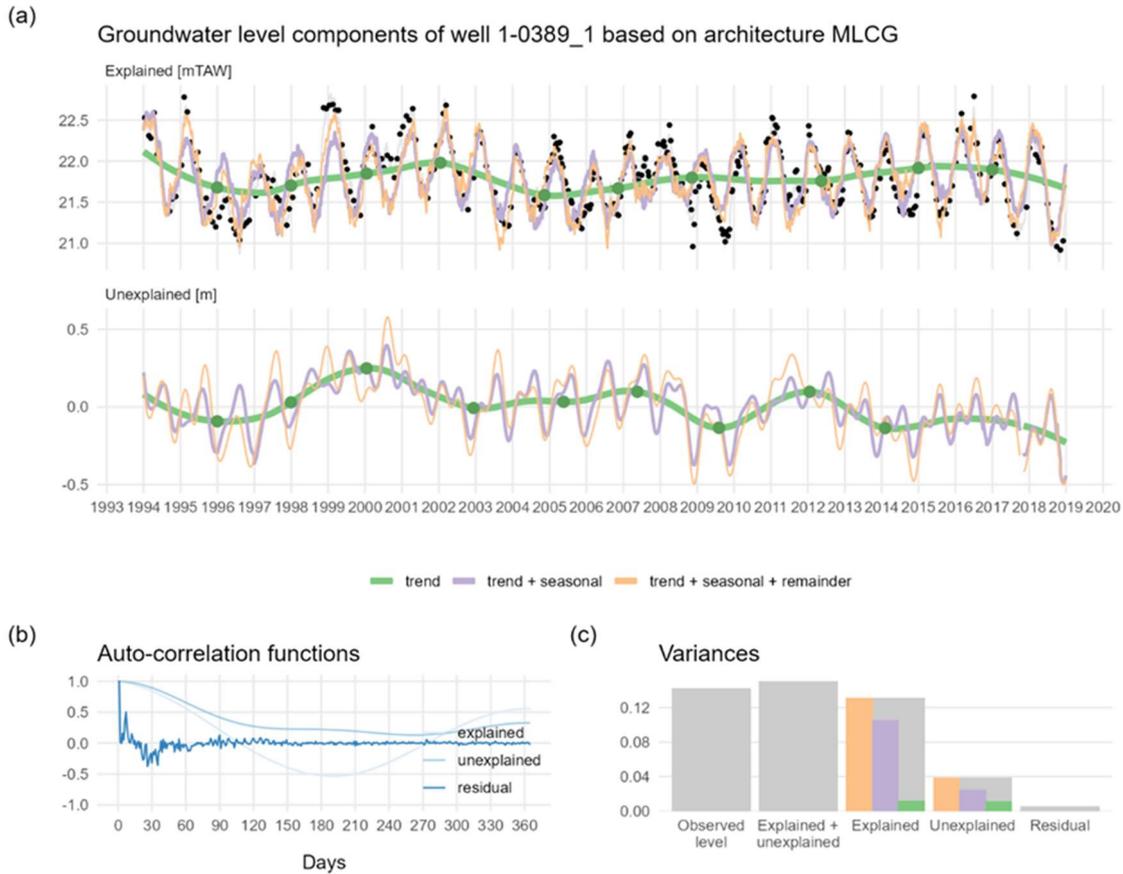


Figure 4.10 Results for the MLCG model architecture, for example well 1-0389\_1.

The MLCG model architecture results are displayed in Figure 4.10. The introduction of a Gaussian process model does seem to increase the explained variance in this case, while again removing the residual auto-correlation. The unexplained part however does show some seasonality, with a rather unexpected phase shift (lower values in winter, larger in summer).

The MLTN model architecture results are displayed in Figure 4.11. By introducing the time-variance here, the variance of the unexplained part seems to be reduced somewhat, compared to that of the MLCN architecture. The difference is however minor. This is similar to the difference between the PLTN and PLCN architectures.

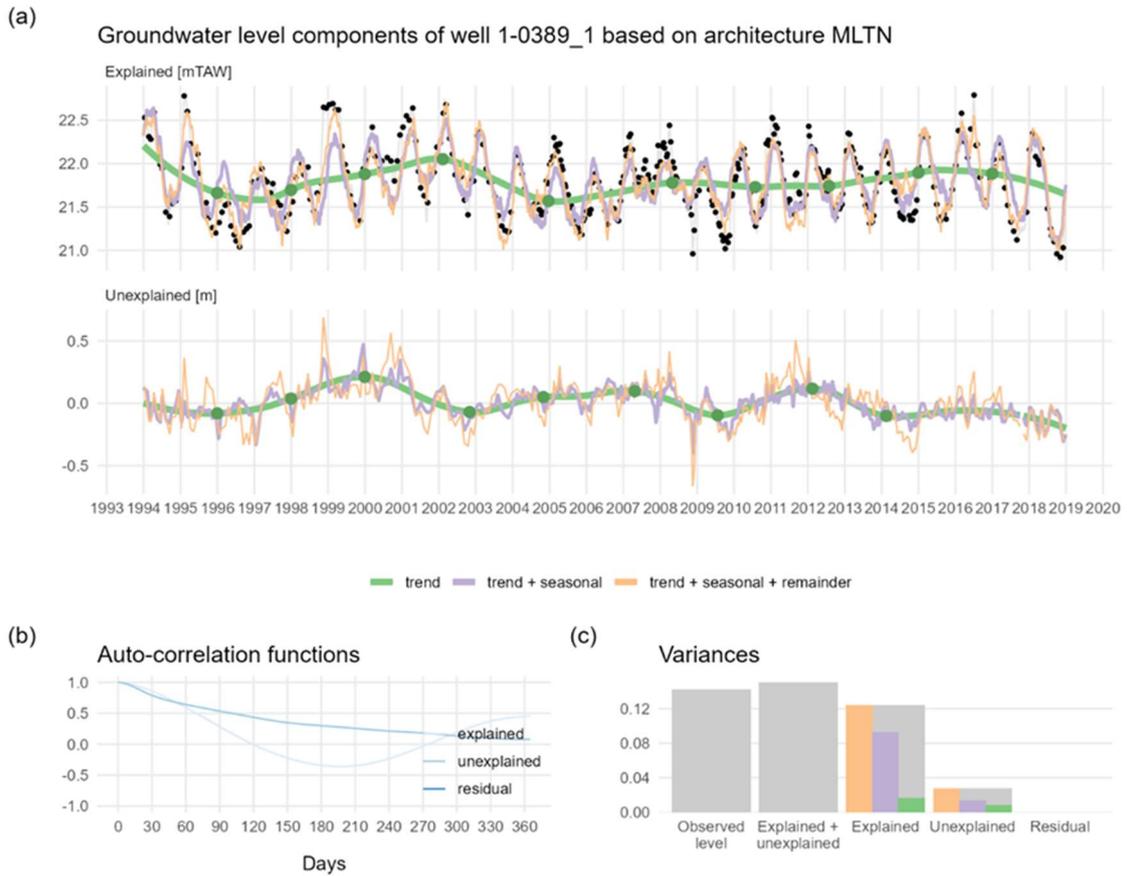


Figure 4.11 Results for the MLTN model architecture, for example well 1-0389\_1.

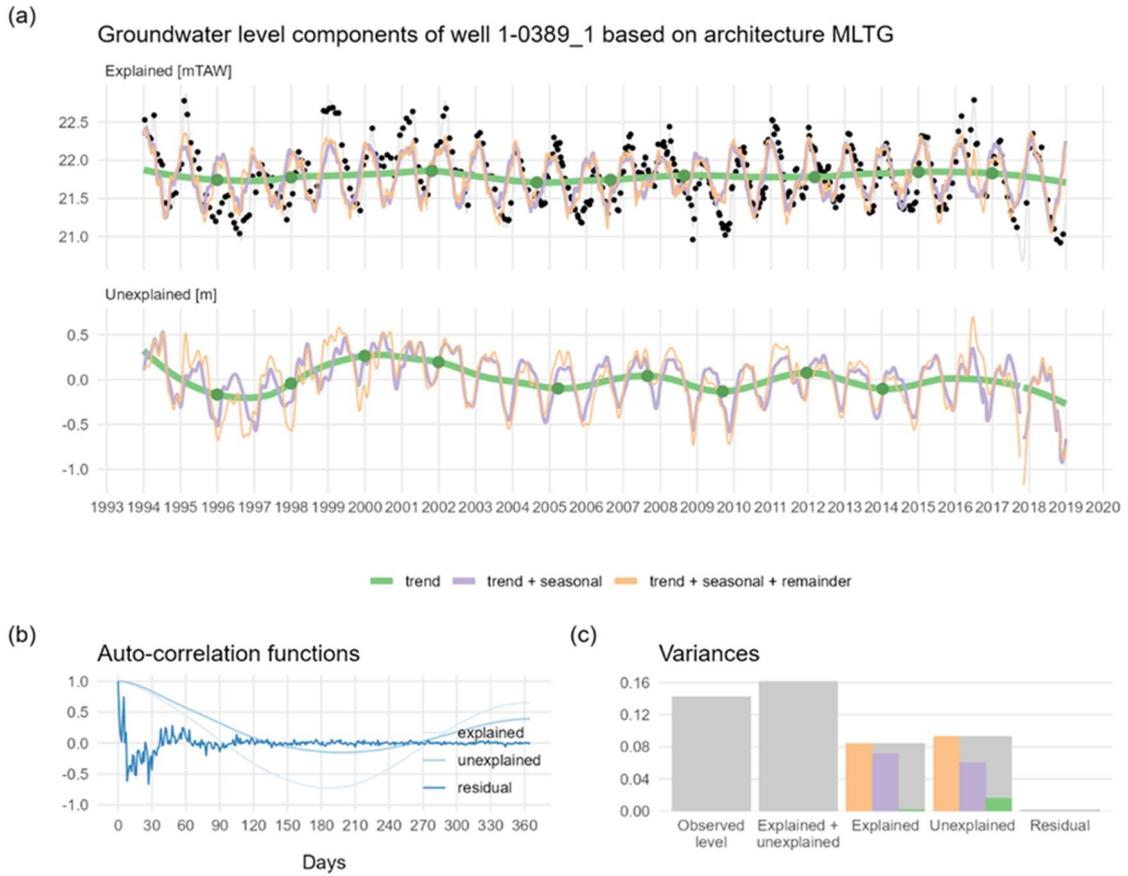


Figure 4.12 Results for the MLTG model architecture, for example well 1-0389\_1.

he MLTG model architecture results are displayed in Figure 4.12. When combining the time-variance and the Gaussian process model, in this case, performance seems to deteriorate again, resulting in a larger variance for the unexplained part, than that for the explained part.

#### 4.1.1.2 Comparison of results from different model architectures

While the above figures provide some feeling of the behaviour of the different individual model architectures, it is rather difficult to compare them. Moreover, we only looked at a single well, which may not be very representative of the performance of the different architectures on the complete set of 21 indicator wells. Hence, we try to illustrate here how the variances of the explained and unexplained parts are related, across the different architectures and wells, as well as the STL features derived from the time series analysis, i.e. the seasonal and trend strengths.

Figure 4.13 provides an idea on how the variances of the explained and unexplained parts relate to each other, across the different model architectures, and displaying the full set of 21 wells. The MLCN and MLTN architectures seem to provide the largest explained and smallest unexplained variances, closely followed by the PLCN and PLTN architectures. The MLTG and PLTG models seem to behave similarly in certain cases, but there is a large spread, and several data points exhibit unexplained variance larger than the explained one. The RLCG and RLTG models clearly perform badly in a systematic way, as most of the data points show much larger unexplained than explained variance.

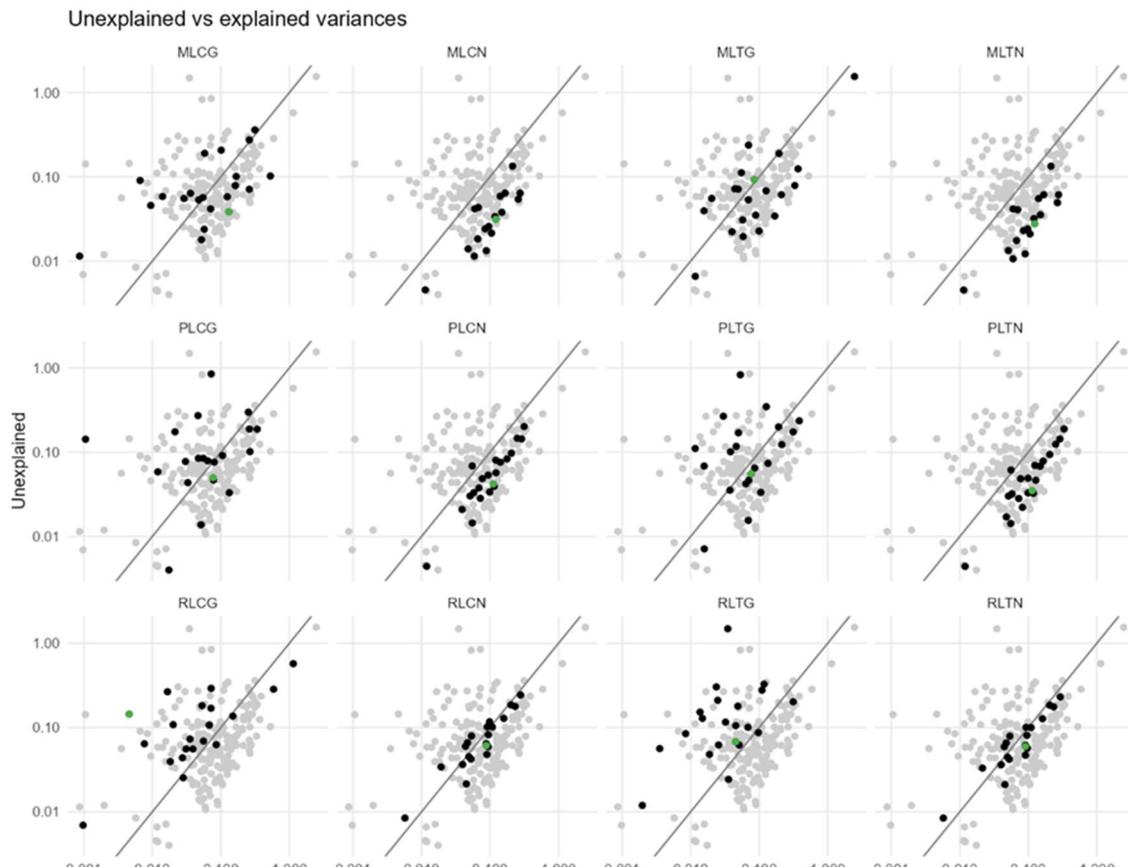


Figure 4.13 Unexplained vs. explained variances per architecture; data points in grey; architecture data points in black; example well in green).



The variances for the example well are shown separately in Figure 4.14. These basically confirm these observations.

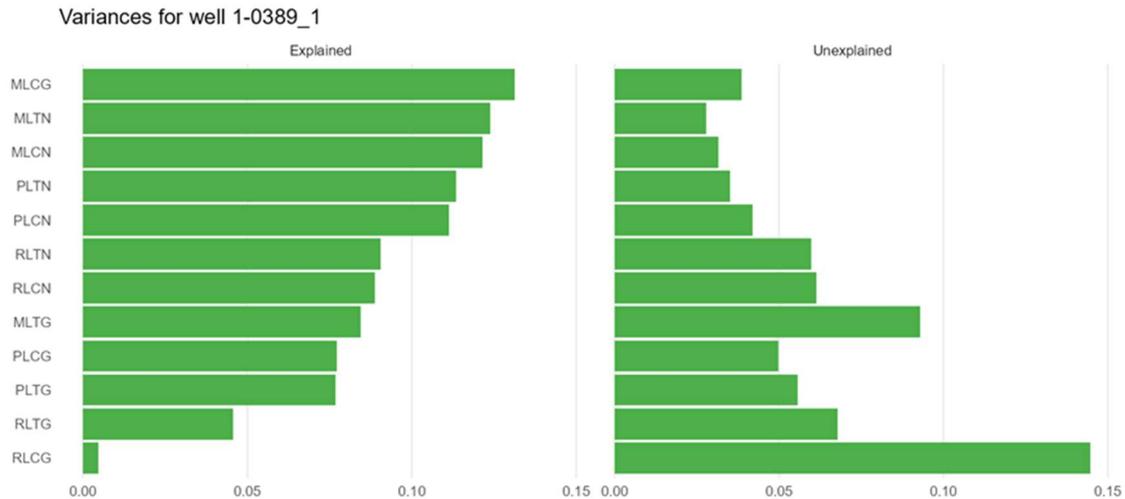


Figure 4.14 Variances for the example well 1-0389\_1, for the explained and unexplained parts, in function of the different model architectures

The variances themselves do however not relate directly to the model performance. The unexplained and explained parts can also be positively or negatively correlated, making interpretation more difficult. To relate the variance of the explained part with the actual performance of the different architectures in terms of the explained part fitting the observed groundwater levels, we provide a scatterplot of the Nash–Sutcliffe efficiency versus the explained variance in Figure 4.15. This does reveal that the MLTN architecture is performing best, closely followed by the MLCN one, and with somewhat systematically lower NSE values, the PLCN and PLTN architectures. Also, it is clear that larger explained variance does not necessarily mean better performance.

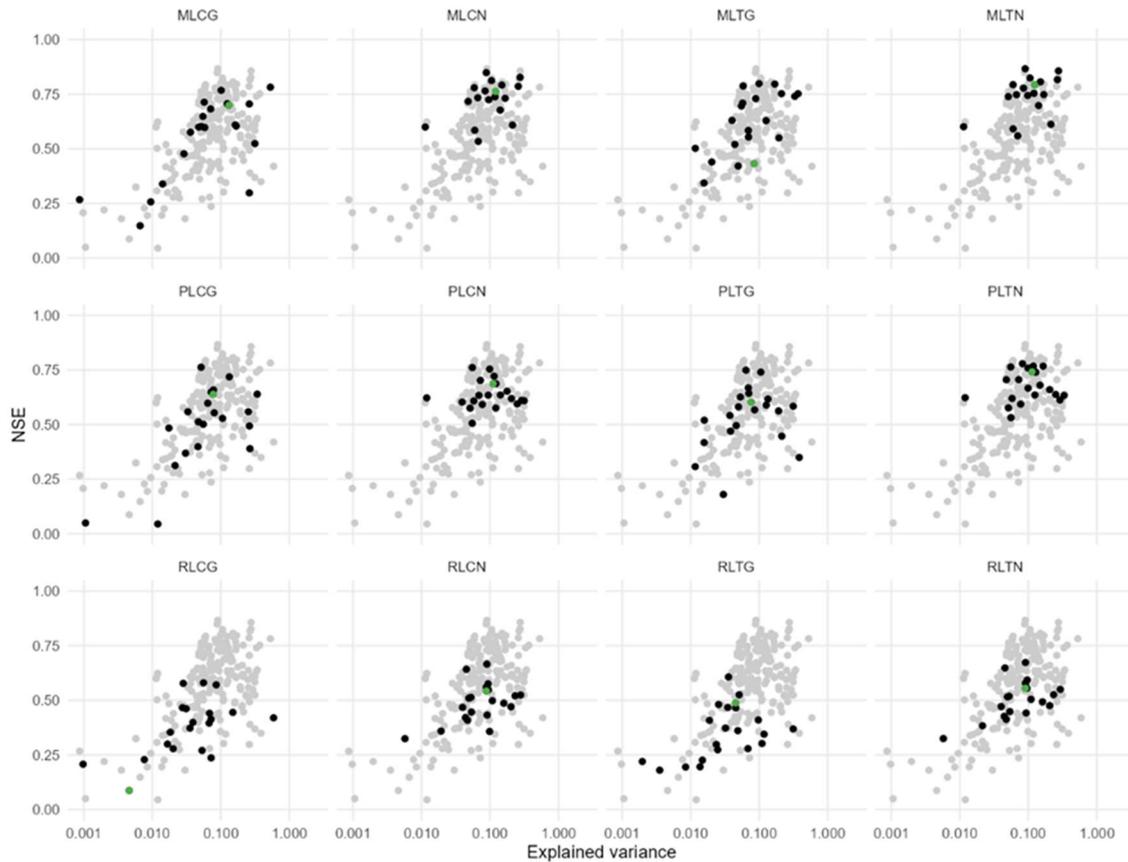


Figure 4.15 Nash–Sutcliffe efficiency (NSE) versus the explained variances per architecture; grey: all data points; black: individual architecture data points; green: example well.

Figure 4.16 and Figure 4.17 provide a similar idea on the seasonal and trend component strengths. The MLCN and MLTN architectures show low unexplained and high explained seasonality, while MLTG is close, and certain data points of the PLCN and PLTN architectures are very similar as well. The architectures using the selected recharge time series clearly do not succeed in achieving large explained seasonality. Trend strengths vary more, and are typically higher for the principal component based architectures, both for the explained and unexplained parts.

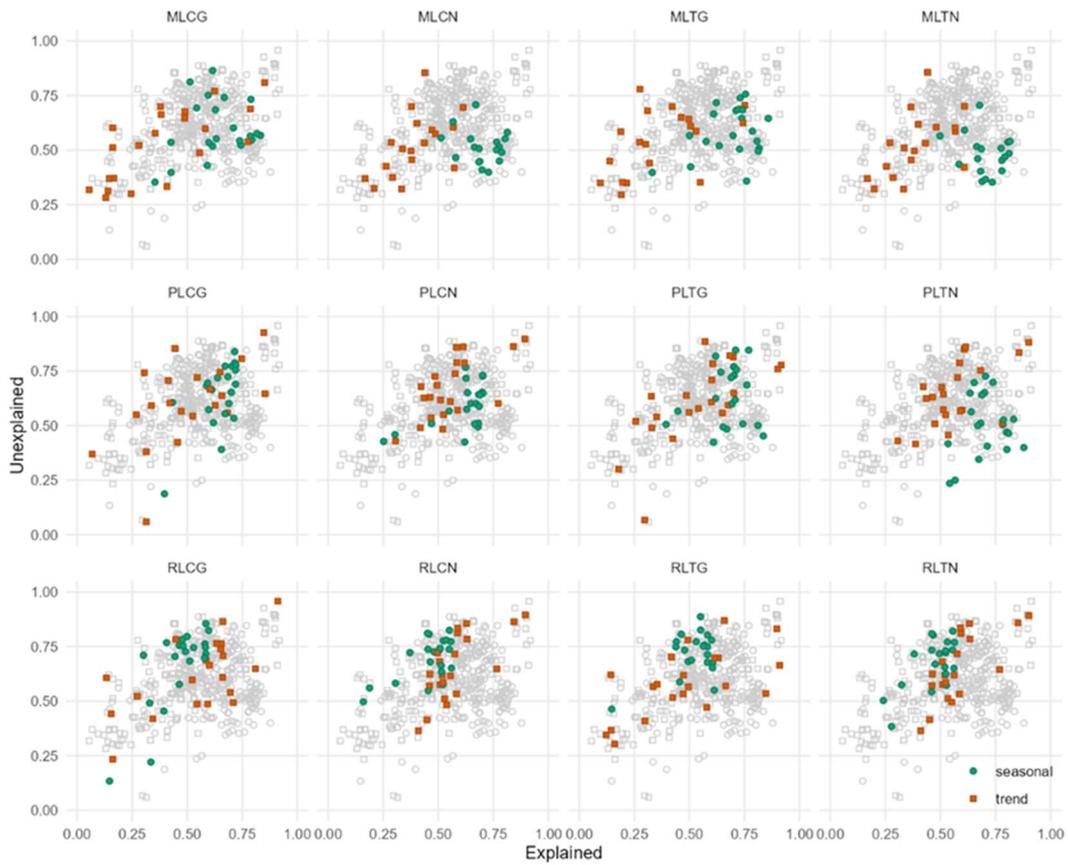


Figure 4.16 Unexplained vs. explained seasonal and trend component strengths per architecture; gray: all data points; green: seasonal component strength; red: strength of trend.

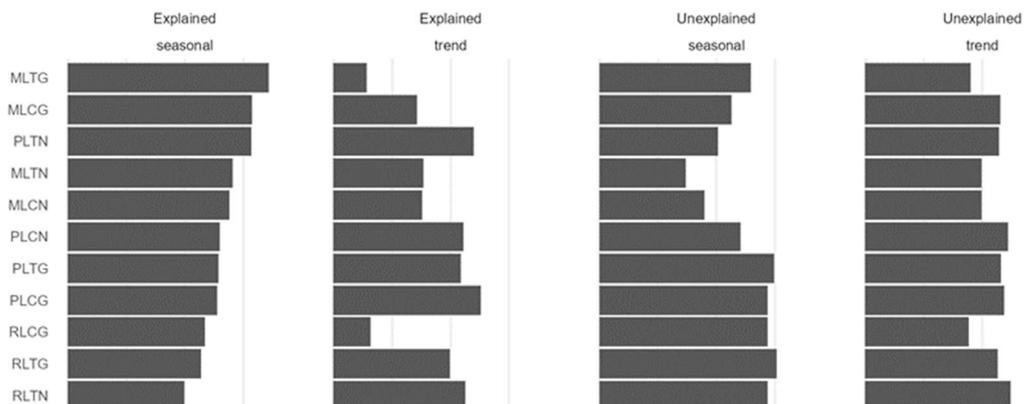


Figure 4.17 Seasonal and trend component strengths in the explained and unexplained parts per architecture for example well 1-0389\_1.

Figure 4.18 allows comparing the different explained and unexplained trend components to each other. Except for the RLCG architecture, most of the trends (both the explained and unexplained) are rather consistent within one of the three groups using different input data sets. Differences between the architectures using a single recharge time series, and those using the principal components also seem to be minor, suggesting the trend estimates are in fact rather robust with respect to the used models. The differences with the meteorological data-based architectures are larger, however, but this may at least in part be due to the different meteorological time series used for the process modelling. The same figures are available for all other indicator wells in appendix C.

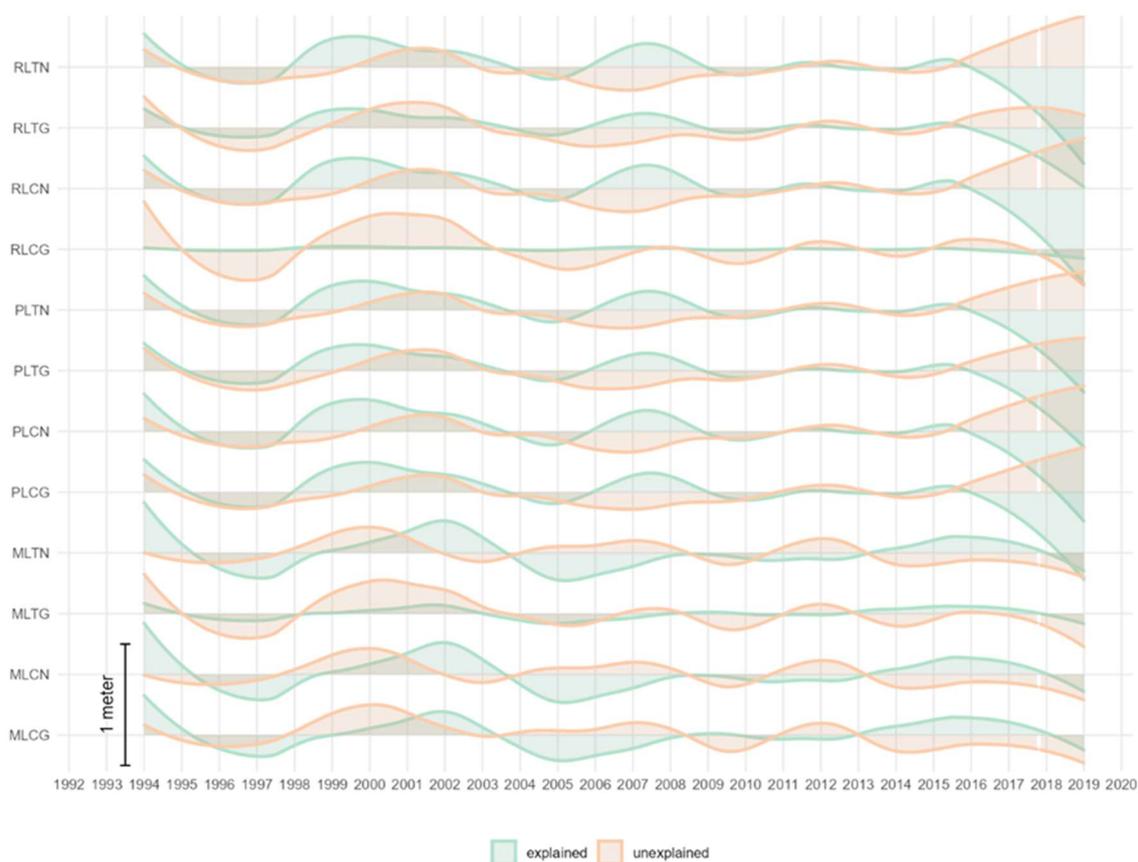


Figure 4.18 Explained and unexplained trend components per architecture for example well 1-0389\_1.

#### 4.1.1.3 Conclusions from IRF analysis of Belgian series

While many of the results remain relatively unexplored at this point (e.g. fitted impulse-response functions (IRF), comparison of recharge estimates from the architectures using meteorological input data with the process model-based recharge), an important conclusion from this exercise is that it seems the trend estimation, which is ultimately the goal, seems to be rather robust, and does not vary a lot over the different model



architectures that were tested. The main differences that we see are those between the architectures using meteorological time series as input and those using the process model recharge database, but these may, at least in part, be related to different meteorological data sources that were used.

Furthermore, the introduction of Gaussian processes to explicitly model the unexplained part of the time series seemed to effectively remove the residual auto-correlation. In many cases, however, the model performance was affected, and we would recommend modifying the approach in future, by allowing larger short-term variability, through the introduction of e.g. a random component. Moreover, attempting to introduce regularization with respect to the seasonal component of the unexplained part, or correlation with the explained part may be interesting as well.

The use of a recharge database seems promising as well, but a more fair comparison starting from the same meteorological time series is required for drawing real conclusions. The principal component approach proved to be useful, but does not allow for feedback of the simulated water table depth on the used recharge. A simple two-way coupling could be explored in future, where for instance a first groundwater level estimate assuming an average depth to the groundwater table, could be used for updating the input recharge times series. Especially for very shallow water tables this could improve the results considerably.

Concerning the investigated architectures, non-linear estimates of recharge other than those based on process modelling, like those by Long (2015) and Collenteur et al. (2021), could be explored as well in future, to make this exercise more complete.

#### **4.1.2 Germany**

No further results are produced for North Rhine-Westphalia but the data are being interpreted in section 4.2.

#### **4.1.3 Netherlands**

The data from Zaadnoordijk and Lourens (2019) have been used to create maps and sections with results of Metran models. Figure 4.20 shows the median precipitation response time  $t_{50}$  (days) for the piezometers with good Metran models in the upper regional aquifer as defined in the groundwater model of the national hydrological model of the Netherlands Hydrological Instrument (<http://nhi.nu/>)

Within the H3O-PLUS project no further analyses have been performed specifically for the Netherlands. The Metran results for the H3O-PLUS head database are presented in Section 4.2.1.

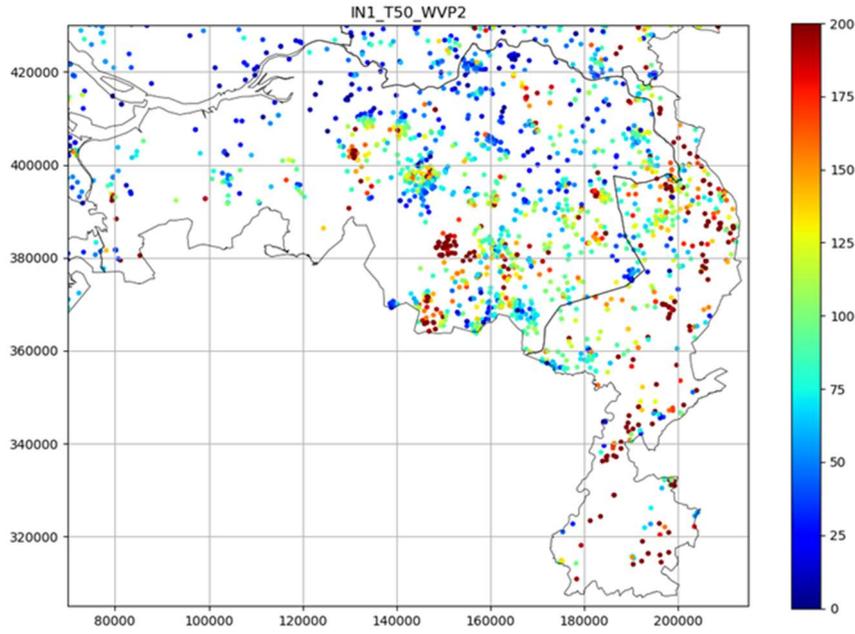


Figure 4.20 map of median precipitation response time (days) for the upper regional aquifer (as defined in the national hydrological model NHI-LHM).

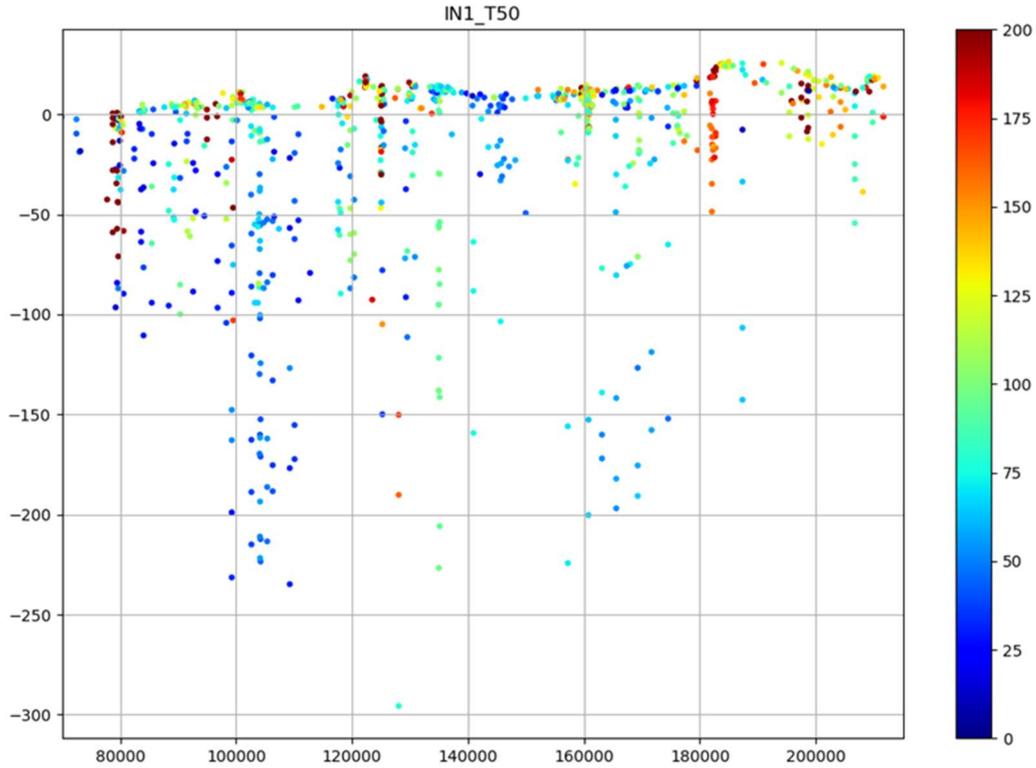


Figure 4.19 section around Y=392500 with median precipitation response time (days) from good Metran models.



## 4.2 Analysis for entire H3O-PLUS area

The selected time series from Belgium, Germany and the Netherlands have been simulated with Metran. The Metran models give the relation between the groundwater heads and the meteorological input: precipitation and either evaporation or temperature. Appendix B shows that Metran determines a very similar the precipitation response with the use of evaporation or temperature. Also, the approximation of the observed heads is quite similar. Therefore, the Metran results for Germany (based on precipitation and temperature) can be shown together with those for Belgium and the Netherlands (based on precipitation and evaporation).

Most of the figures in this section are screen captures from the *H3O-Plus web viewer* <https://dev.grondwatertools.nl/gwsinbeeld/login>. It has been created based on the Dutch tool for spatial and temporal patterns of groundwater heads <http://www.grondwaterstandeninbeeld.nl> which uses the heads from the national subsurface database <https://www.dinoloket.nl/en/subsurface-data>.

The tool allows selections based on:

- Period of the measurements;
- Depth range or Geological Formation of the piezometer.

For each piezometer, properties of the time series model or trends of the measured heads for different periods can be indicated.

Topographic maps and aerial imagery are available as background in plan view. Figure 4.21 shows an example with a selection based on a geological Formation with the median response time of the precipitation response that has been determined in the Metran models for the head time series in the piezometers within the Sterksel Formation. The background is the aerial imagery.

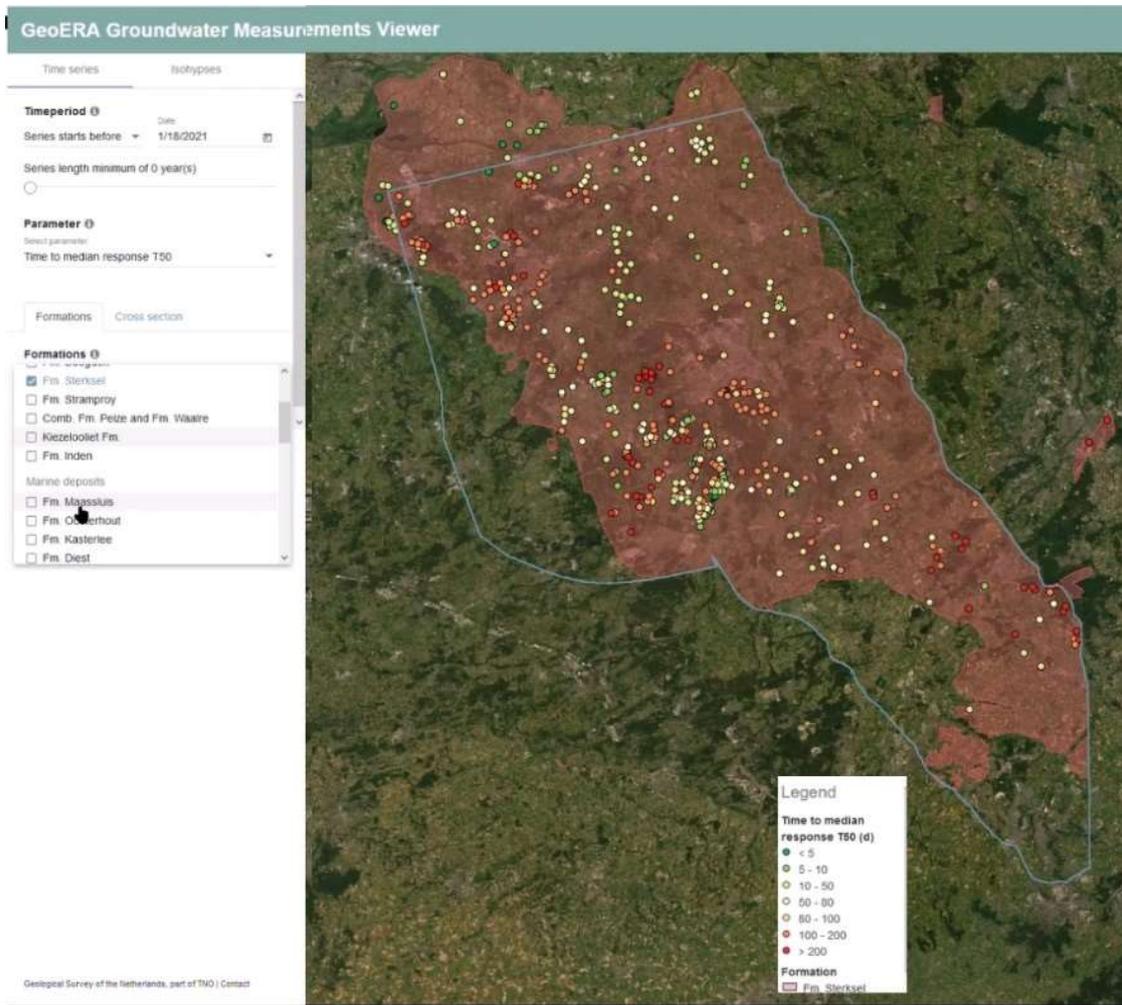


Figure 4.21: Median response time of precipitation for piezometers in the Sterksel Formation from the GeoERA Groundwater Measurements Viewer.

Cross sections are shown with the geological formations as background and can be annotated additionally with groundwater abstractions, land use, groundwater age zones, and a brackish-saline interface (Figure 4.22). This figure shows the vertical head differences between the subsequent piezometers in multi-piezometer monitoring wells. The coloured bars are drawn between the filter bottom of a piezometer and the top of the next deeper one. The colour indicates the flow direction: red downward, blue upward and yellow neutral. There are no multi-level wells in the Southeast (the right hand side of the cross section) to show the downward flow here. Moving to the Northwest, the blue bars indicate the upward flow toward the river Meuse. Further to the Northwest, there is upward flow from the Kiezoloöliet Formation (brown colours) to the Waalre Formation (dark yellow and orange) as well as downward flow from the overlying Sterksel Formation (reddish colours) and Bortel Formation (bright yellow) down to the Waalre Formation, mostly related to the groundwater extractions in the Waalre Formation.

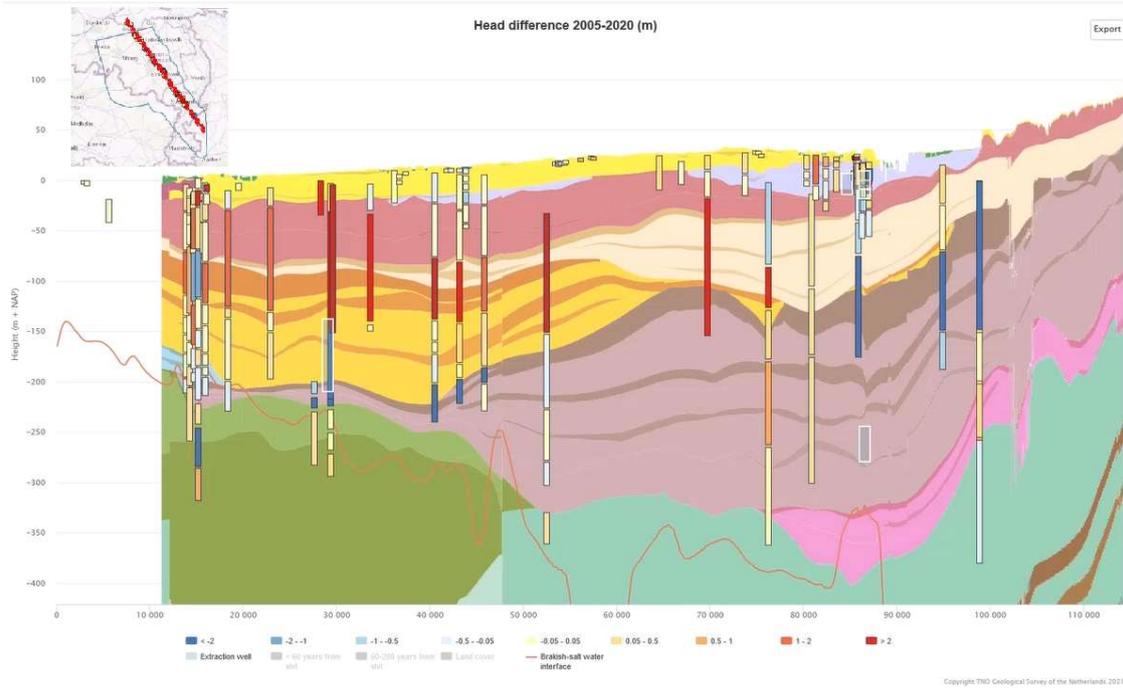


Figure 4.22: Cross section with head differences between subsequent piezometers in multi-level monitoring wells from the GeoERA Groundwater Measurements Viewer.

#### 4.2.1 Metran results for entire H30-PLUS area

The GeoERA Viewer (<https://www.grondwatertools.nl/gwsinbeeld/login>) provides access to the H30-PLUS groundwater head database and the Metran models (Figure 4.23).

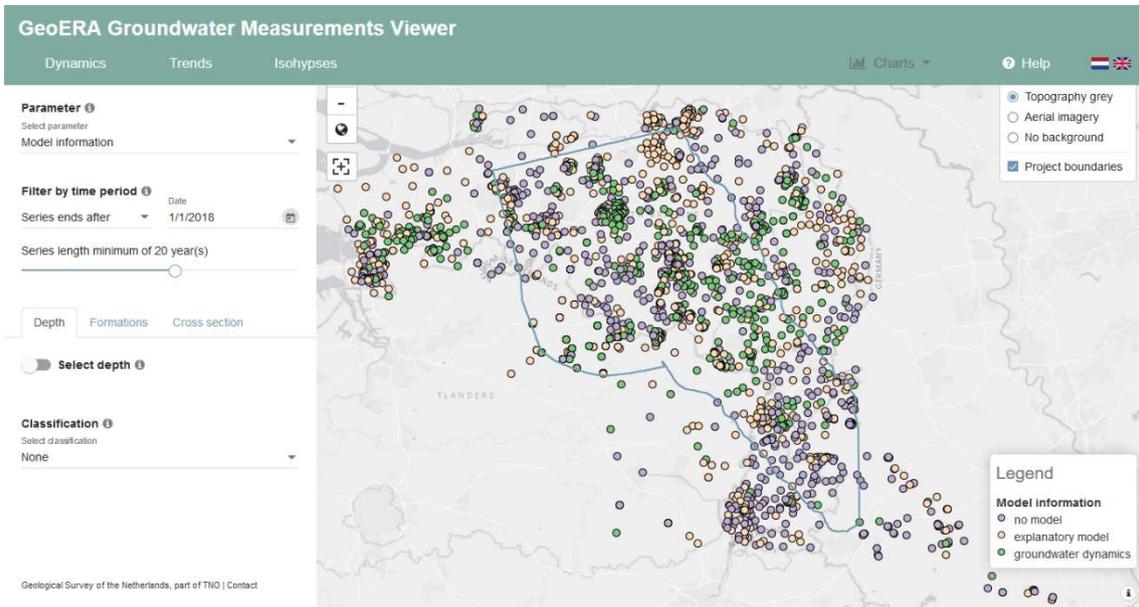


Figure 4.23 Observation wells in H3O-PLUS database with quality of time series model for upper piezometer

The viewer allows selections based on depth or connected geological formation of the piezometers and the period of available measurements.

Figure 4.24 shows the piezometers for various depths with an indication of the quality of the Metran model (grey: bad model – ‘no model’, yellow: reasonable model – ‘explanatory model’, green: good model – ‘groundwater dynamics’).

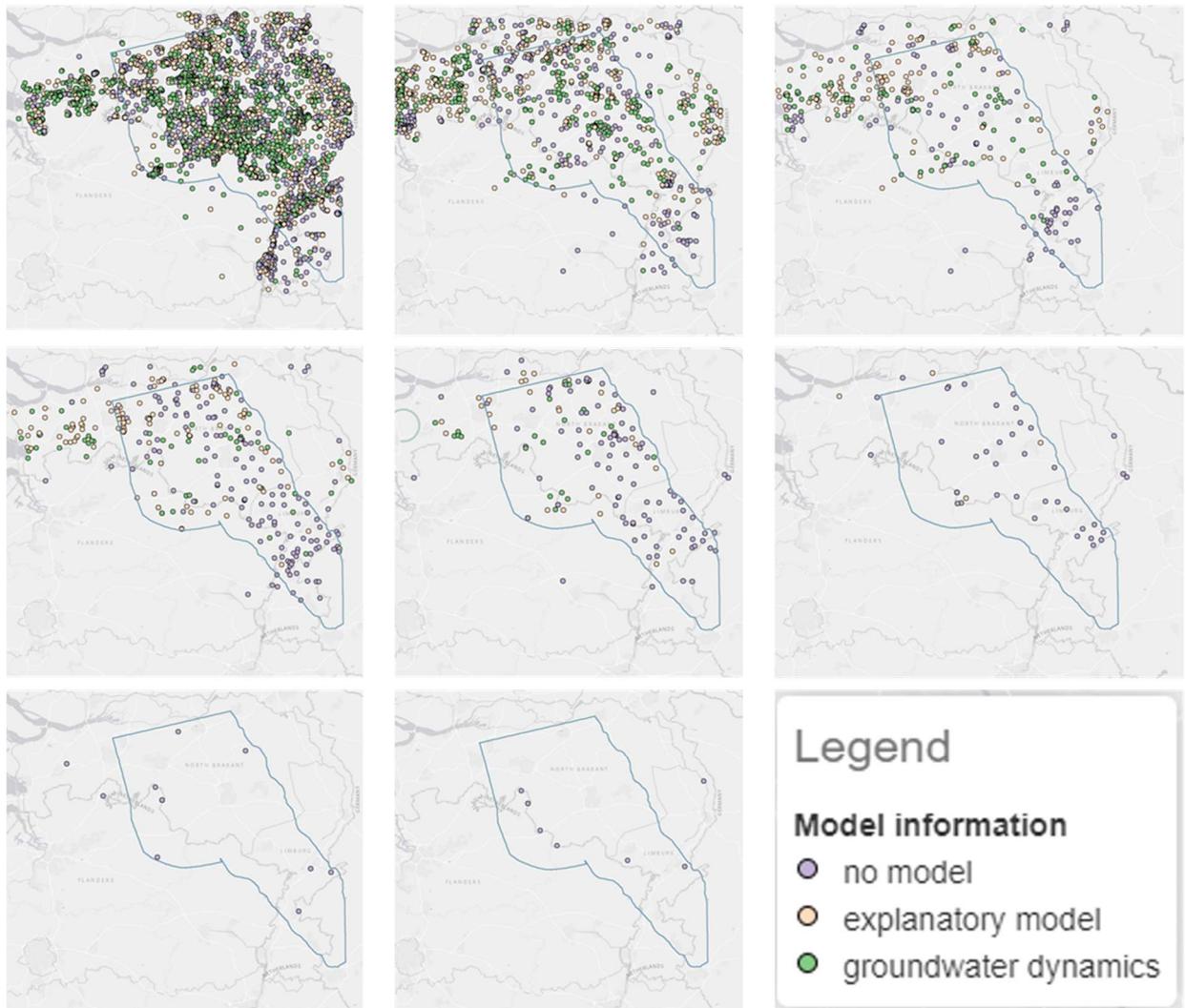


Figure 4.24 Model quality per depth range (top row: +40 to 0mNAP, -1 to -49, -51 to -99; middle row: -100 to -200, -200 to -300, -300 to -400; bottom row: -400 to -500; -460 to -540 mNAP).

The number of piezometers decreases with depth (see Figure 4.24). Also the number of successful time series models decrease with depth. For practically all piezometers below -250 m NAP, no time series model could be derived.

Figure 4.25, Figure 4.26, and Figure 4.27 show characteristics of the precipitation response in a long cross section from the North to the Southeast of the H30-PLUS area.

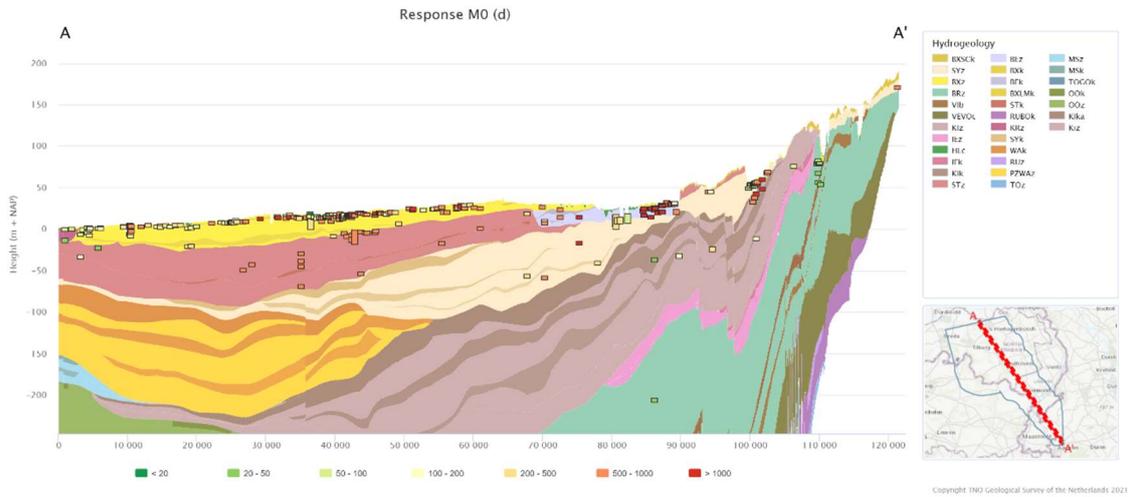


Figure 4.25 Total precipitation response M0 in Northwest-Southeast cross section.

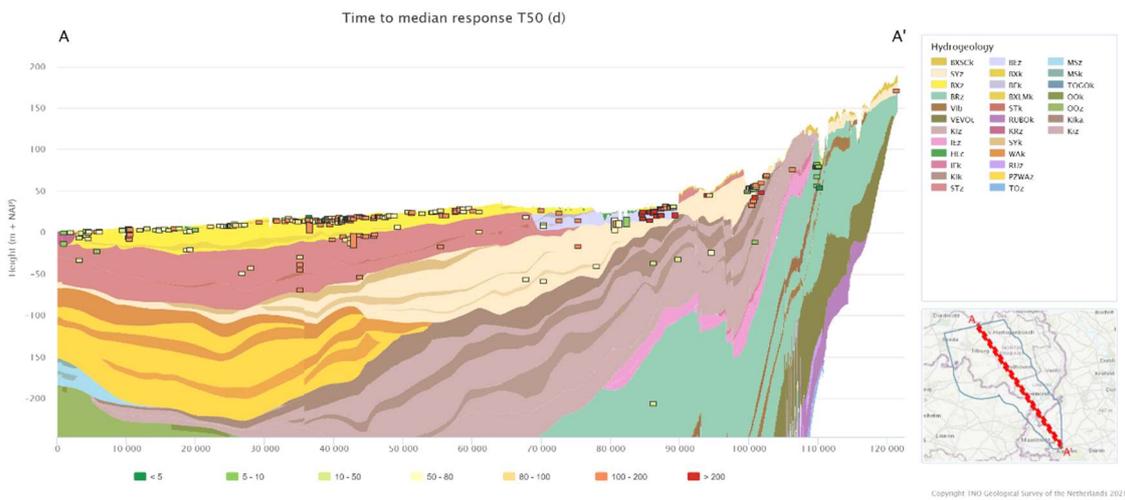


Figure 4.26 Median response time t50 in Northwest-Southeast cross section.

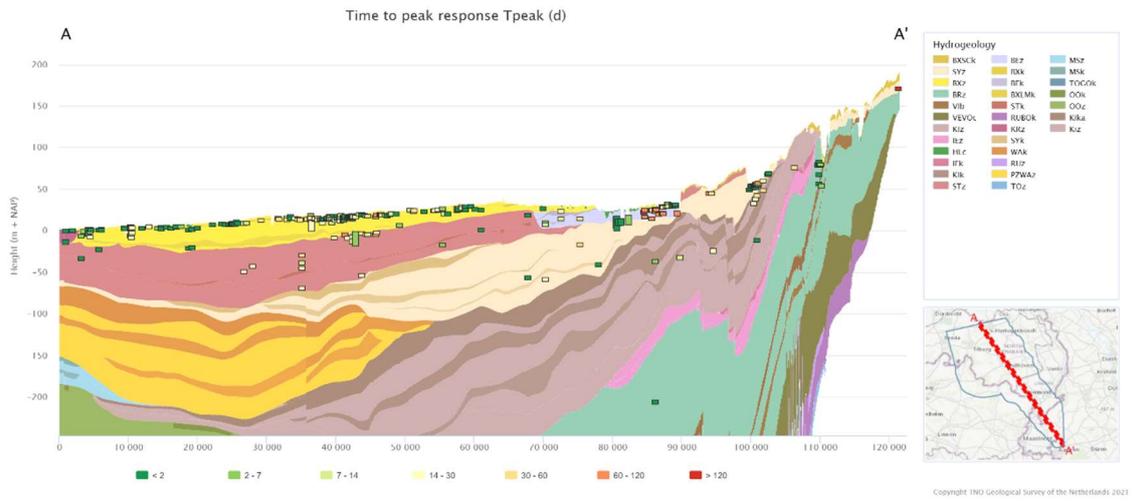


Figure 4.27 Peak response time in Northwest-Southeast cross section.

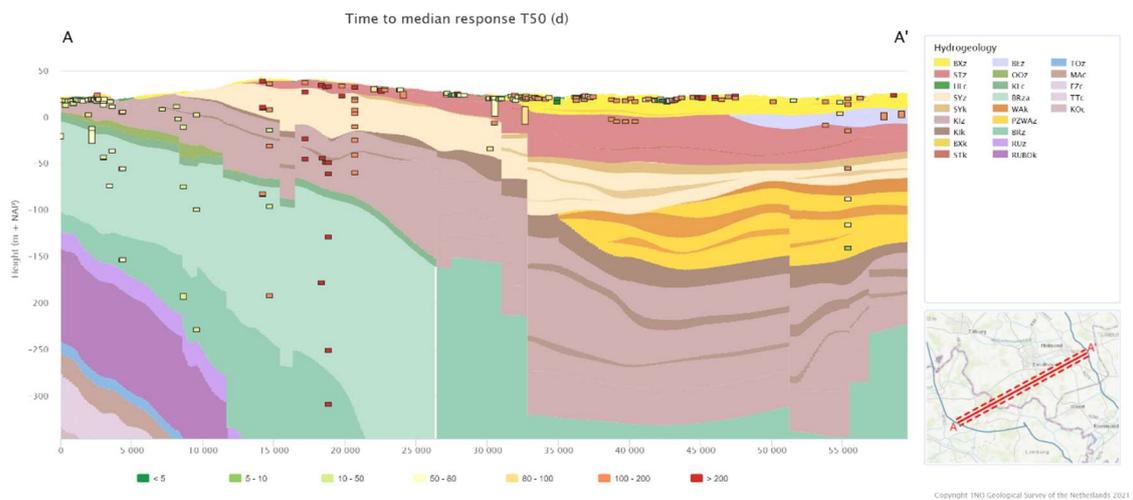


Figure 4.28 Median response time in Northerly West-East cross section.

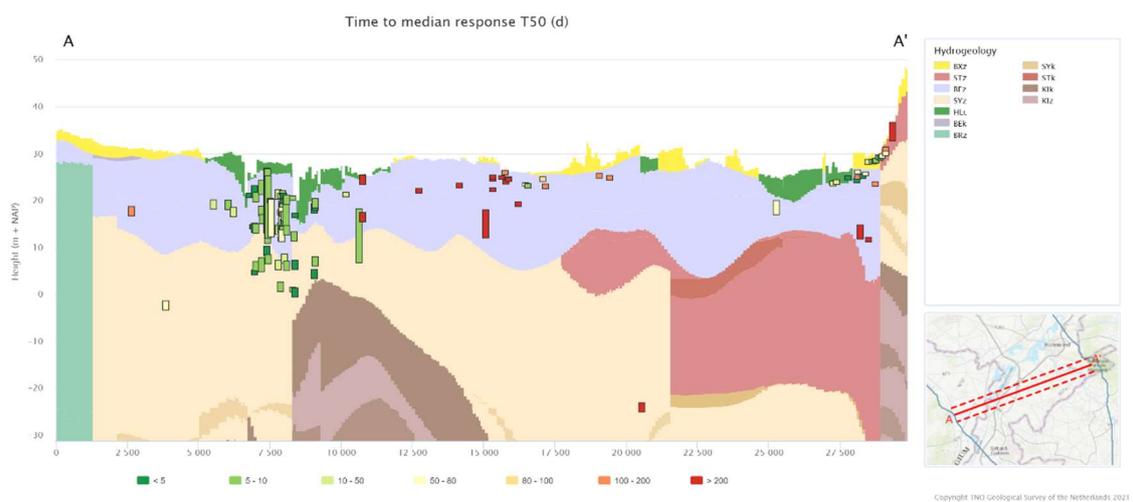


Figure 4.29 Median response time in Southerly West-East cross section.

Figure 4.28 and Figure 4.29 show the median precipitation response time in two cross sections in West-East direction

#### 4.2.2 Trend analysis of groundwater head time series

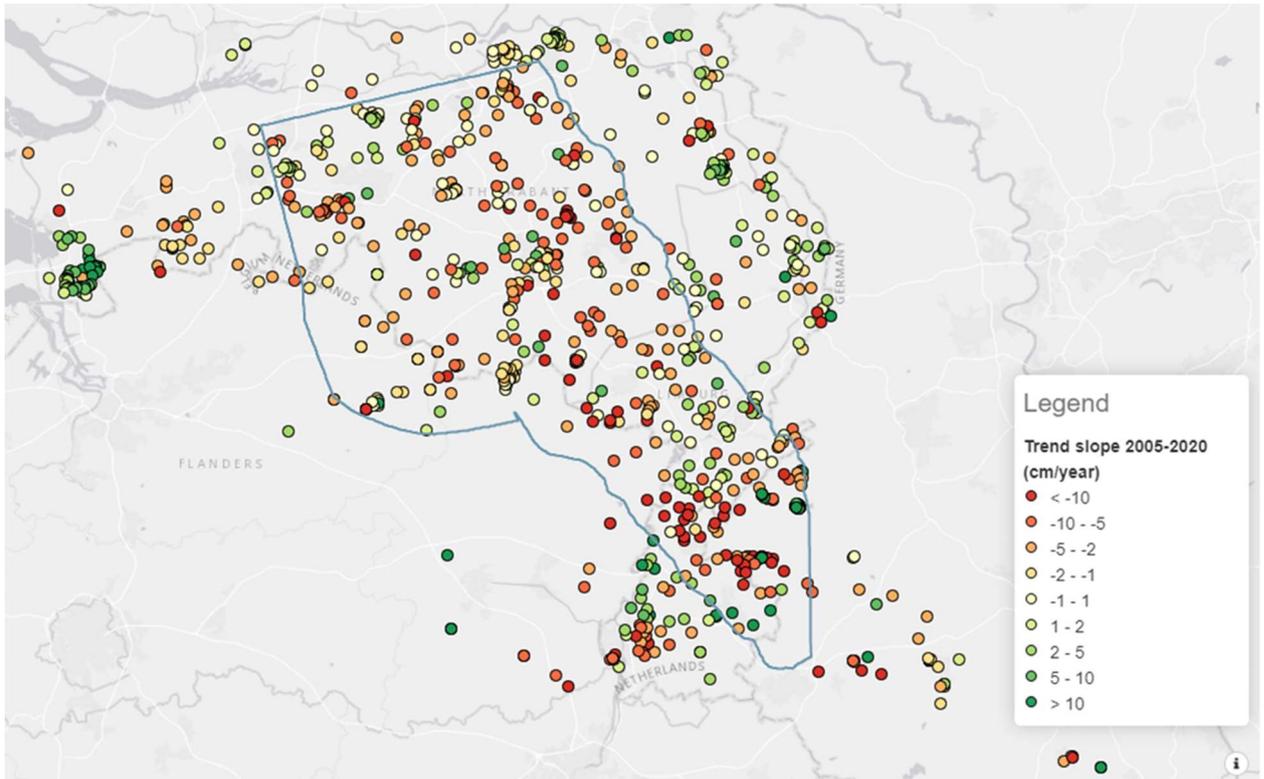


Figure 4.30 Trend groundwater heads 2005-2020

In addition to the Metran models, the groundwater head series themselves have been tested for trends. The Mann-Kendal test has been used for the existence of a significant trend. If a significant trend exists, it was quantified using the Sen procedure, which is insensitive to outliers.

Figure 4.30 shows the trend in the groundwater heads for the period from the year 2005 to the year 2020. Many point have a negative trend. For the period from the year 1995 to the year 2010, the green colours of a positive trend dominate (Figure 4.31).

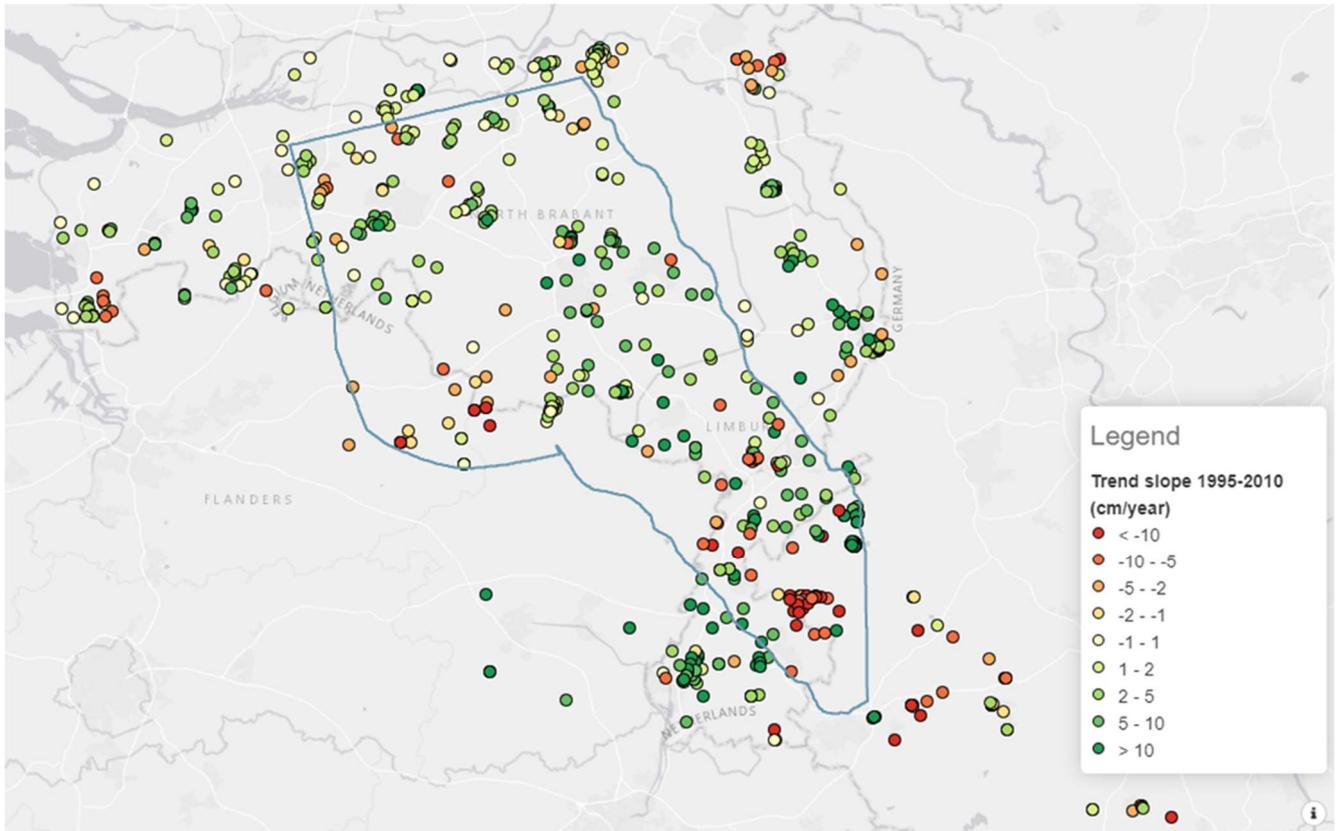


Figure 4.31 Trend groundwater heads 1995-2010.

In order to analyse the differences between the two periods and to determine to what extent the trends are caused by changes in precipitation and evaporation or by other influences, Metran models are investigated for selected piezometers (see subsection 4.2.3).

#### 4.2.3 Separation of trends into meteorological and other influences

The time series for a number of monitoring wells are selected to investigate the trends that were determined in subsection 0 together with the timeseries model results presented in 4.2.1.

The selected monitoring wells are: 7-0355 (with 4 piezometers, located in Belgium), B51E0078 (with 10 piezometers, located in the Netherlands) and three single piezometer wells in Germany: \_210401114, \_010300387, \_010404211.

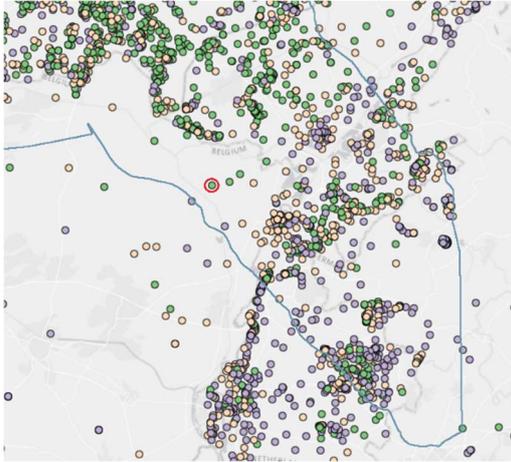


Figure 4.32 Location of Belgian monitoring well 7-0355 (red circle).

Figure 4.32 shows the location of the Belgian monitoring well 7-0355. The elevations of the piezometers and the geological formations are given in Table 4.1.

Table 4.1 Elevation of piezometers of Belgian monitoring well 7-0355.

Piezometer	Top (m+NAP)	Bottom (m+NAP)	Formation
1	18.61	16.61	Beegden
2	-31.39	-33.39	Stramproy
3	-141.39	-143.39	Kiezoloöliet
4	-201.39	-203.39	Kiezoloöliet

The timelines for the four piezometers are shown in Figure 4.33. The two upper and the two lower ones are relatively close.

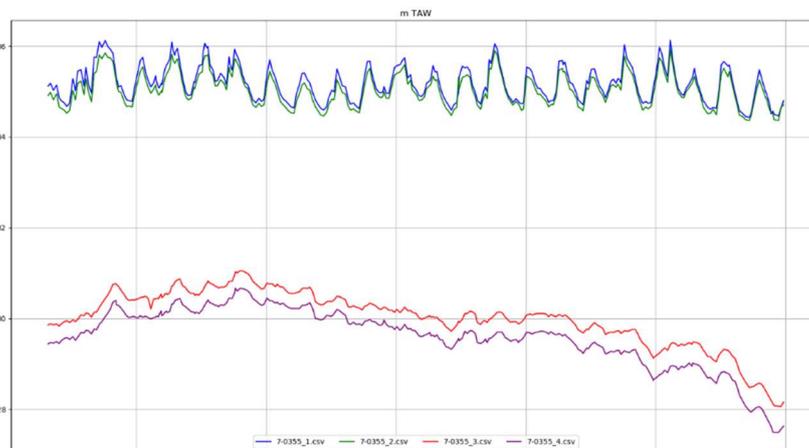


Figure 4.33 Timeseries for the four piezometers of Belgian monitoring well 7-0355.

The head gradient is consistently downward indicating vertical downward flow from the surface to the depth of the deepest piezometer. The seasonal fluctuation is stronger in the timelines of the two upper. The Metran models for these two are good. The simulation and the residue (difference between simulation and measurements) suggests that there is downward trend related to precipitation minus evaporation and that the anthropogenic influences do not cause a trend (Figure 4.34).

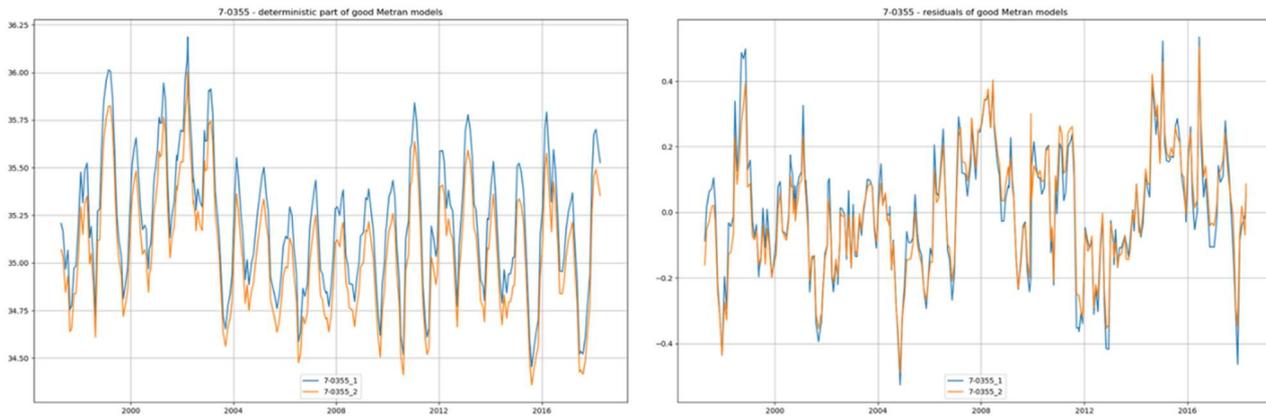


Figure 4.34 deterministic part and residus of the good Metran models for Belgian well 7-0355 (piezometer 1 and 2).

Metran did not give a reliable model for the third and fourth piezometer. However, the rise around the year 2000 and the decline at the end in these series (Figure 4.33) suggest that these series also contain a major response to the precipitation and evaporation (compare left graph in Figure 4.34).

The timelines from a Dutch well in Figure 4.35 show a similar rise around 2000 and decline at the end.

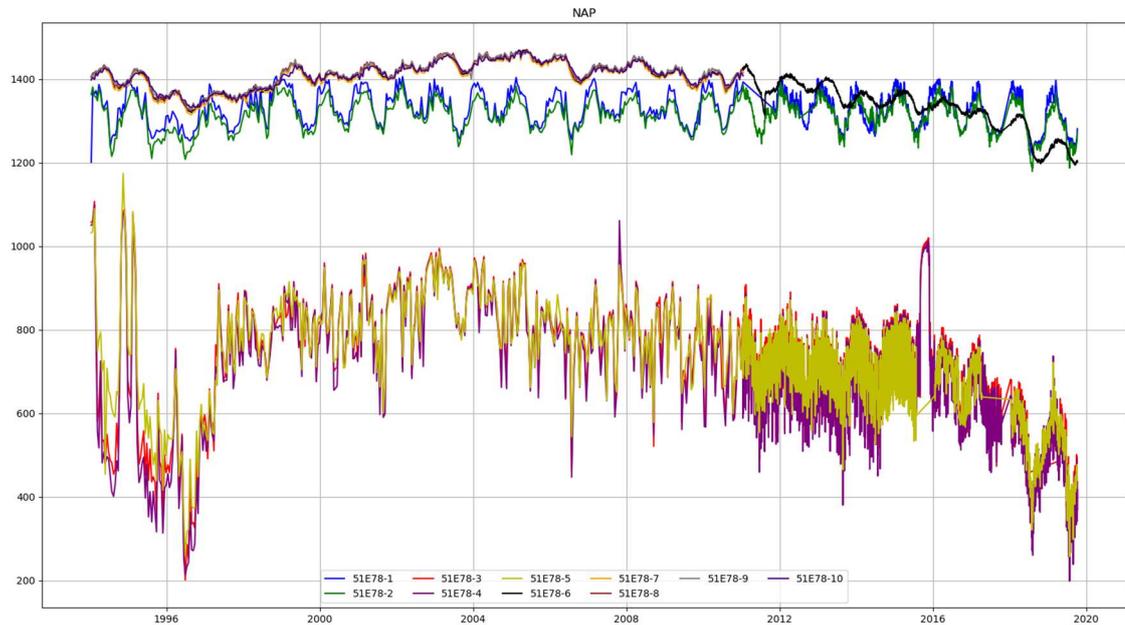


Figure 4.35 timeseries for piezometers of Dutch monitoring well B51E0078.

The three German timeseries in Figure 4.36 have quite different behaviours.

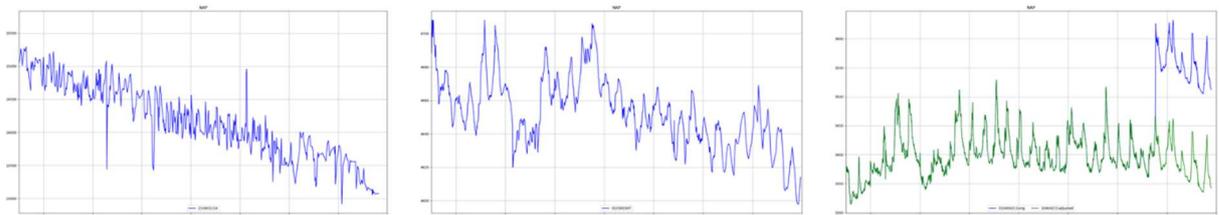


Figure 4.36 Time series for the German piezometers 210401114 (left), 010300387 (centre), and 010404211 (right; original series in blue, corrected series in green).

For 010404211 a positive trend for 2005-2020 had been determined, while most trends for this period are negative (Figure 4.30) due at least in part to the precipitation and evaporation (e.g. Figure 4.34). However, this positive trend seems to be caused by an error in the height of the last part of the series. When this is corrected (green line in Figure 4.36), there no longer is a positive trend.

## 5 CONCLUSIONS

Groundwater head data within the provinces Noord-Brabant and Limburg from the Dutch national subsurface data repository (<https://www.DINOloket.nl>), Flemish data from VMM and SCK, and German data within North Rhine-Westphalia from the Erftverband have been added, together with precipitation, evaporation, and temperature data from the Dutch, Belgian, and German Meteorological Institutes.

The report represents the first transboundary visualisation of groundwater heads and cross-border patterns of groundwater depletion for the study region based on the respective monitoring networks in the participating countries. These visualisations are now available through a dedicated transboundary webviewer; the GeoERA Groundwater Head Web viewer allows the visualization of groundwater head monitoring with time series and derived data like response time to precipitation and trends (Figure 5.1).

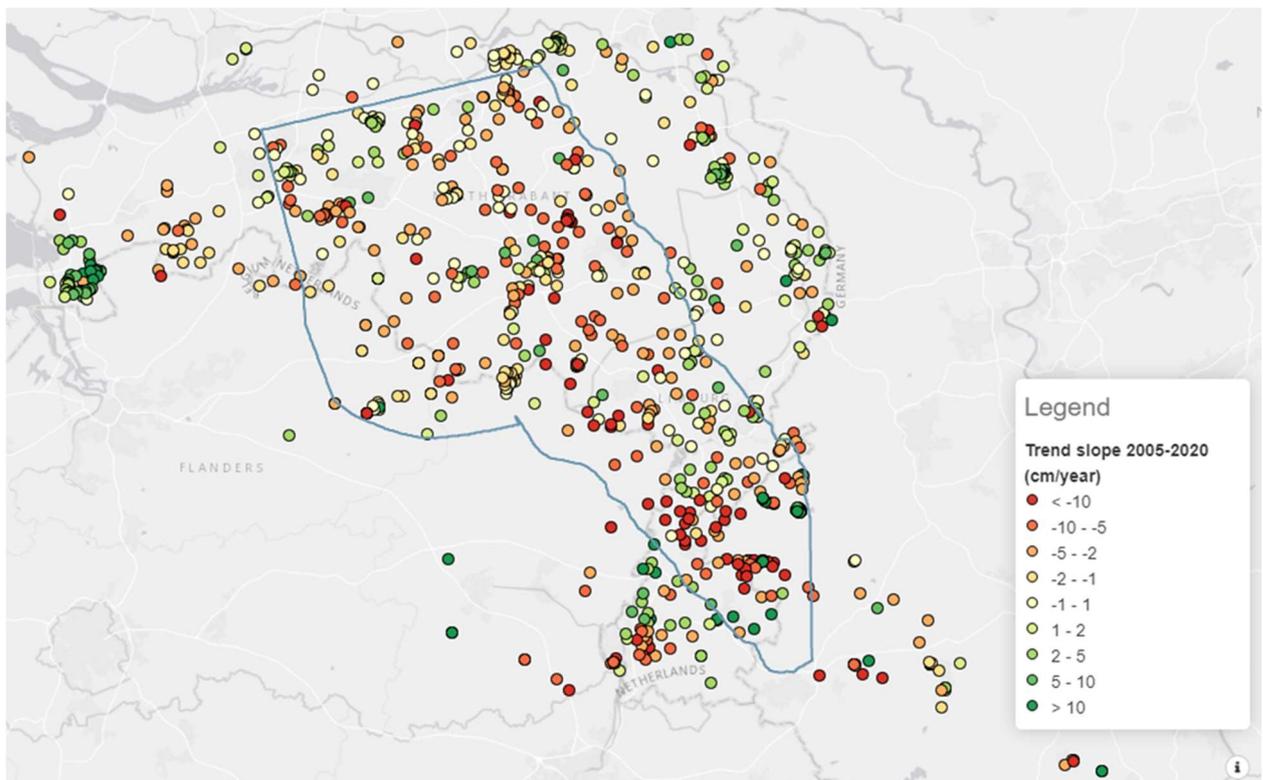


Figure 5.1 Cross boundary visualization of trend in groundwater heads (period 2005-2020). The data can also be visualized in cross-sections through the whole H3O PLUS area rendering a 3D visualization tool for the areas.

The results of the current report will be input for the next deliverable (D3.6) where we will explore the current protection and management strategies for shallow resources and deep paleo resources in the region. For this aim, the consortium cooperates with the regional stakeholders involved in the project, using the newly developed information and visualization system to help groundwater managers to harmonize groundwater strategies.



## 6 REFERENCES

Allen, R. G. (1992). Standard Reference Evapotranspiration Calculations: REF-ET. In *Irrigation and Drainage: Saving a Threatened Resource—In Search of Solutions* (pp. 140-145). ASCE.

Bengtsson, H. (2020). A Unifying Framework for Parallel and Distributed Processing in R using Futures, arXiv:2008.00553, 2020

Berendrecht, W.L., F.C. van Geer (2016) A dynamic factor modeling framework for analyzing multiple groundwater head series simultaneously, *Journal of Hydrology*, volume 536, pages 50-60, <http://dx.doi.org/10.1016/j.jhydrol.2016.02.028>

Besbes, M, G. de Marsily (1984) From Infiltration to Recharge: use of a parametric transfer function, *Journal of Hydrology*, volume 74, pages 271-293.

Collenteur, R. A., Bakker, M., Klammler, G., & Birk, S. (2021). Estimation of groundwater recharge from groundwater levels using nonlinear transfer function noise models and comparison to lysimeter data. *Hydrology and Earth System Sciences*, 25(5), 2931-2949.

Cooley, D. (2020). sfheaders: Converts Between R Objects and Simple Feature Objects. R package version 0.4.0. <https://CRAN.R-project.org/package=sfheaders>

Cronshey, R. (1986). *Urban hydrology for small watersheds* (No. 55). US Department of Agriculture, Soil Conservation Service, Engineering Division.

de Vries, A. & Luraschi, J. (2020). nomnoml: Sassy 'UML' Diagrams. R package version 0.2.3. <https://CRAN.R-project.org/package=nomnoml>

Di Ciacca, A. (2020). Spatially distributed recharge and groundwater – surface water interactions in groundwater models: from the field to the catchment scale. KUL - Katholieke Universiteit Leuven.

Golding, N. (2019). greta: simple and scalable statistical modelling in R. *Journal of Open Source Software*, 4(40), 1601. URL <http://dx.doi.org/10.21105/joss.01601>.

Golding, N. (2020). greta.gp: Gaussian Process Modelling in greta. R package version 0.1.5.9001.

Hafen, R. (2016). stlplus: Enhanced Seasonal Decomposition of Time Series by Loess. R package version 0.5.1. <https://CRAN.R-project.org/package=stlplus>

Hester, J. & Wickham, H. (2020). fs: Cross-Platform File System Operations Based on 'libuv'. R package version 1.5.0. <https://CRAN.R-project.org/package=fs>



Heuvelmans, G., A. Louwyck, J. Lermytte (2011) Distinguishing between management-induced and climatic trends in phreatic groundwater levels, *Journal of Hydrology*, volume 411, pages 108-119, doi:10.1016/j.jhydrol.2011.09.039.

Hyndman, R.J., & Athanasopoulos, G. (2021) *Forecasting: principles and practice*, 3rd edition, OTexts: Melbourne, Australia. OTexts.com/fpp3. Accessed on 2021-07-12.

Kroes, J.G., Van Dam, J.C., Groenendijk, P., Hendriks, C.F.A. & Jacobs, C.M.J. (2008) SWAP version 3.2 Theory description and user manual. Alterra-report 1649. WUR, Wageningen, Netherlands.

Leterme, B., Mallants, D., & Jacques, D. (2012). Sensitivity of groundwater recharge using climatic analogues and HYDRUS-1D. *Hydrology and Earth System Sciences*, 16(8), 2485-2497.

Leunk, I., A. van Doorn (2017) Trendanalyse grondwaterstands- en stijghoogtegegevens Maasstroomgebied (2012-2016) (Trend analysis groundwater table en head data Meuse river basin (2012-2016); in Dutch), report KWR 2017.046, commissioned by provinces of Noord-Brabant and Limburg, KWR, Nieuwegein, Netherlands.

Long, A. J. (2015). RRAWFLOW: Rainfall-response aquifer and watershed flow model (v1. 15). *Geoscientific Model Development*, 8(3), 865-880.

Lu, M. (in preparation) Quantifying and predicting river-aquifer exchange in lowland floodplains (Nete, Demer and Dijle). Ongoing PhD project, SCK CEN & KU Leuven.

Lu, M., Rogiers, B., Beerten, K., Gedeon, M., Huysmans, M. (in preparation). Exploring river-aquifer interactions and hydrological system response using baseflow separation, impulse response modelling and time series analysis in three temperate lowland catchments.

Müller, K. (2020). here: A Simpler Way to Find Your Files. R package version 1.0.1. <https://CRAN.R-project.org/package=here>

Neuwirth, E. (2014). RColorBrewer: ColorBrewer Palettes. R package version 1.1-2. <https://CRAN.R-project.org/package=RColorBrewer>

Obergfell, C., M. Bakker, K. Maas (2019) Estimation of average diffuse aquifer recharge using time series modeling of groundwater heads. *Water Resources Research*, 55(3), 2194-2210, <https://doi.org/10.1029/2018WR024235>.

Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal* 10 (1), 439-446, <https://doi.org/10.32614/RJ-2018-009>

Pedersen, T. L. (2020). patchwork: The Composer of Plots. R package version 1.1.1. <https://CRAN.R-project.org/package=patchwork>



Pedersen, T. L. & Shemanarev, M. (2021). ragg: Graphic Devices Based on AGG. R package version 1.1.3. <https://CRAN.R-project.org/package=ragg>

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Richardson, N., Cook, I., Keane, J., François, R., Ooms, J. & Apache Arrow (2021). arrow: Integration to 'Apache' 'Arrow'. R package version 4.0.1. <https://CRAN.R-project.org/package=arrow>

Roberts, S., Osborne, M., Ebdon, M., Reece, S., Gibson, N., & Aigrain, S. (2013). Gaussian processes for time-series modelling. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1984), 20110550.

Rogiers, B., Vandersteen, K., Labat, S., Beerten, K., Gedeon, M. (2016). Water table mapping at the cAt site. SCK CEN-R-6147/12758018, 70 pp.

Rogiers, B., Jacques, D., Gedeon, M., Labat, S., Vandersteen, K., Beerten, K., Laloy, E. (2018). Daily spatial groundwater table dynamics over a 15-year period from monthly time series of different lengths. *Computational Methods in Water Resources XXII*. June 3 - 7, 2018. Saint-Malo, France.

Simunek, J., Van Genuchten, M. T., & Sejna, M. (2005). The HYDRUS-1D software package for simulating the one-dimensional movement of water, heat, and multiple solutes in variably-saturated media. *University of California-Riverside Research Reports*, 3, 1-240.

Van Orshoven, J., Maes, J., Vereecken, H., Feyen, J., & Dudal, R. (1988). A structured database of Belgian soil profile data. *Pedologie: bulletin van de Belgische bodemkundige vereniging*, 38(2), 191-206.

Von Asmuth, J. R., Bierkens, M. F., & Maas, K. (2002). Transfer function-noise modeling in continuous time using predefined impulse response functions. *Water Resources Research*, 38(12), 23-1.

Von Asmuth, J. R., & Bierkens, M. F. (2005). Modeling irregularly spaced residual series as a continuous stochastic process. *Water Resources Research*, 41(12).

Von Asmuth, J. R., Maas, K., Bakker, M., & Petersen, J. (2008). Modeling time series of ground water head fluctuations subjected to multiple stresses. *Groundwater*, 46(1), 30-40.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D. A., François, R., ... & Yutani, H. (2019). Welcome to the Tidyverse. *Journal of open source software*, 4(43), 1686.



Wilke, C. O. (2021). ggridges: Ridgeline Plots in 'ggplot2'. R package version 0.5.3. <https://CRAN.R-project.org/package=ggridges>

Zaadnoordijk, W.J., S.A.R. Bus, A. Lourens, W.L. Berendrecht (2019) Automated Time Series Modeling for Piezometers in the National Database of the Netherlands, Groundwater, Vol. 57, No. 6, 834–843, doi: 10.1111/gwat.12819.

Zaadnoordijk, W.J., A. Lourens (2019) Groundwater dynamics in the aquifers of the Netherlands, TNO report TNO 2019 R12031, TNO Geological Survey of the Netherlands, Utrecht, Netherlands.

Zeileis, A., Leisch, F., Hornik, K. & Kleiber C. (2002). strucchange: An R Package for Testing for Structural Change in Linear Regression Models. Journal of Statistical Software, 7(2), 1-38. URL <https://www.jstatsoft.org/v07/i02/>

Zeileis, A., Kleiber, C., Kraemer, W. & Hornik, K. (2003). Testing and Dating of Structural Changes in Practice. Computational Statistics & Data Analysis, 44, 109-123.



## APPENDIX A – DATABASE

### Setup database

Collection of files each containing one time series

### Format for time series files

.csv in plain ASCII with:

- Double quotes " " around text strings
- Period . as decimal symbol
- No thousands symbol
- Lengths in meters
- Date in format: dd/mm/yyyy
- Time in format: hh:mm

### Contents of the .csv file:

- Line with header for metadata:
  - ID (string)
  - Quantity ("Head", "Precipitation", "Makkink", "Penman-Monteith", . . .)
  - Unit ("m", maybe "°C" for temperature instead of evaporation)
  - X coordinate
  - Y coordinate
  - Surface level
  - Top piezometer/measurement point elevation
  - Top filter
  - Bottom filter
  - Projection (for X, Y coordinates)
  - Reference level (for heights)
  - Own aquifer code
  - H3Oplus aquifer code
  - . . .
- line with metadata
- Header for time series "Date", "Time", "Value", ? "Qualification" ?
- Lines with Date, Time, Value, Qualification (e.g. "<" if piezometer dry)



## APPENDIX B – EVAPOTRANSPIRATION (ET) VS TEMPERATURE (T)

### B.1 ET vs. T

The temperature models have all been created with the temperature time series for Eindhoven because this is the KNMI weather station that is closest to most of the groundwater head monitoring locations (see Table 0.1).

*Table 0.1 KNMI stations used for evaporation data for models of Dutch groundwater head series with Makkink evaporation.*

count	code station
1294	370 Eindhoven
671	350 Gilze-Rijen
616	380 Maastricht
448	375 Volkel
257	356 Herwijnen
101	391 Arcen
20	377 Eil

The variation of the temperature in the H3Oplus area could cause deviations in the time series models. However, Figure 0.1 shows, that the temperature differences between the most used weather stations of Table 0.1 are quite small. The patterns are so similar that it is unlikely that the temperature variations influence the quality of the time series models.

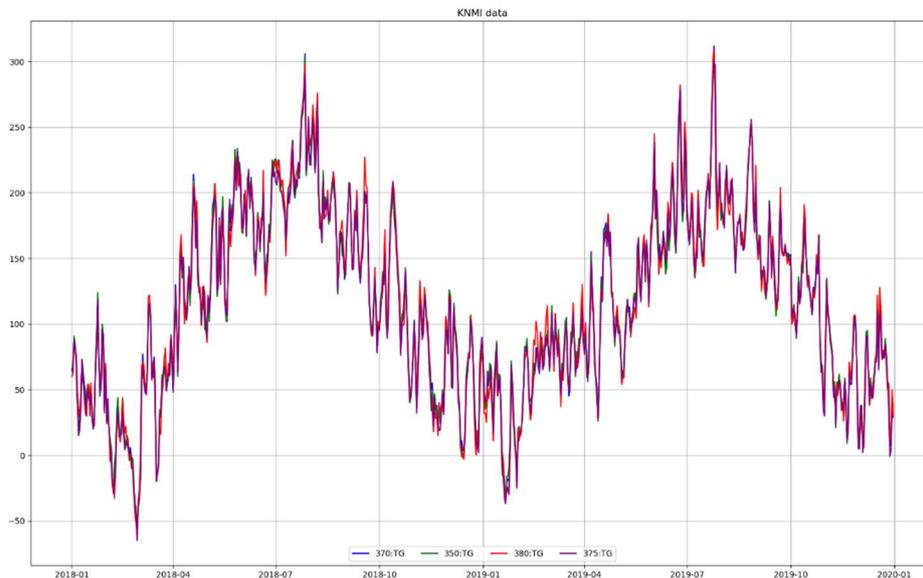


Figure 0.1 Temperature for the KNMI weather stations Eindhoven (370), Gilze-Rijen (350), Maastricht (380), and Volkel (375).

Figure 0.2 shows the temperature and evaporation models for B60B0220001 side by side. These decent models (REGIMEOK) deviate little from the measurements. The root mean squared error (RMSE) is 8.5 and 9.4 cm for evaporation and temperature respectively. The total precipitation responses are respectively 12038 and 10397. The average precipitation response times (51 and 24) deviate much more from the 1:1 line (see section 9.3 of this Appendix B). Still the base elevations of the models are almost the same (3218 and 3220 cm for evaporation and temperature respectively). The difference in the precipitation response does not lead to visible differences in the precipitation contribution. The differences in the evaporation and temperature contributions (Figure 0.2, third row, right and left graph) do reflect the differences between the shape of the evaporation and temperature graph. They only give small differences in the residuals.

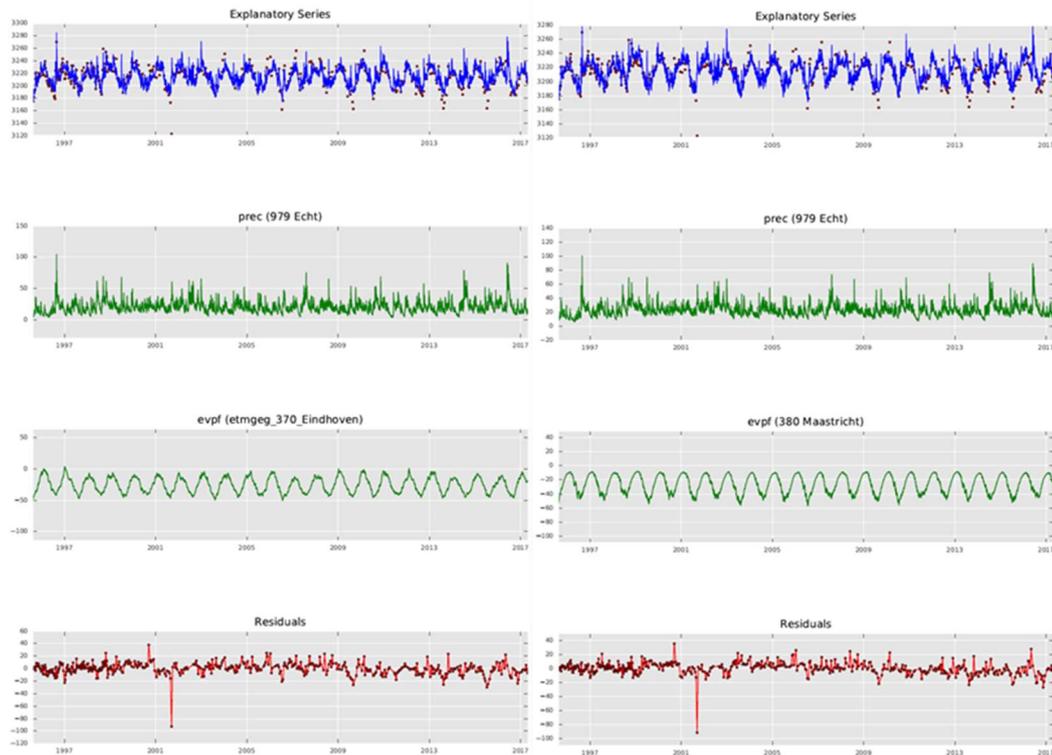


Figure 0.2 Output of model with temperature (left) and with Makkink evaporation (right) for the Dutch piezometer B60B0220001.

Figure 0.3 shows the time series models for B45E0419001, which differ more in the total precipitation response ( $M_0$  is 59752 for the evaporation model and 94179 for the temperature model – the point is away from the 1:1 line as can be seen in section 9.3 of this Appendix B). Together with the differences in average precipitation response times (363 and 550 for evaporation and temperature respectively), this gives a smoother precipitation contribution with a smaller range for the evaporation model. The base elevations of the two models are again almost equal (865 vs. 859 cm) and –surprisingly– also the RMSE (25.2 vs. 25.0 cm). So, for this piezometer the difference in shape of the evaporation and temperature time lines leads to a different precipitation response, and would give more disturbance in the analysis of spatial patterns of precipitation response than B60B0220001 of Figure 0.2.

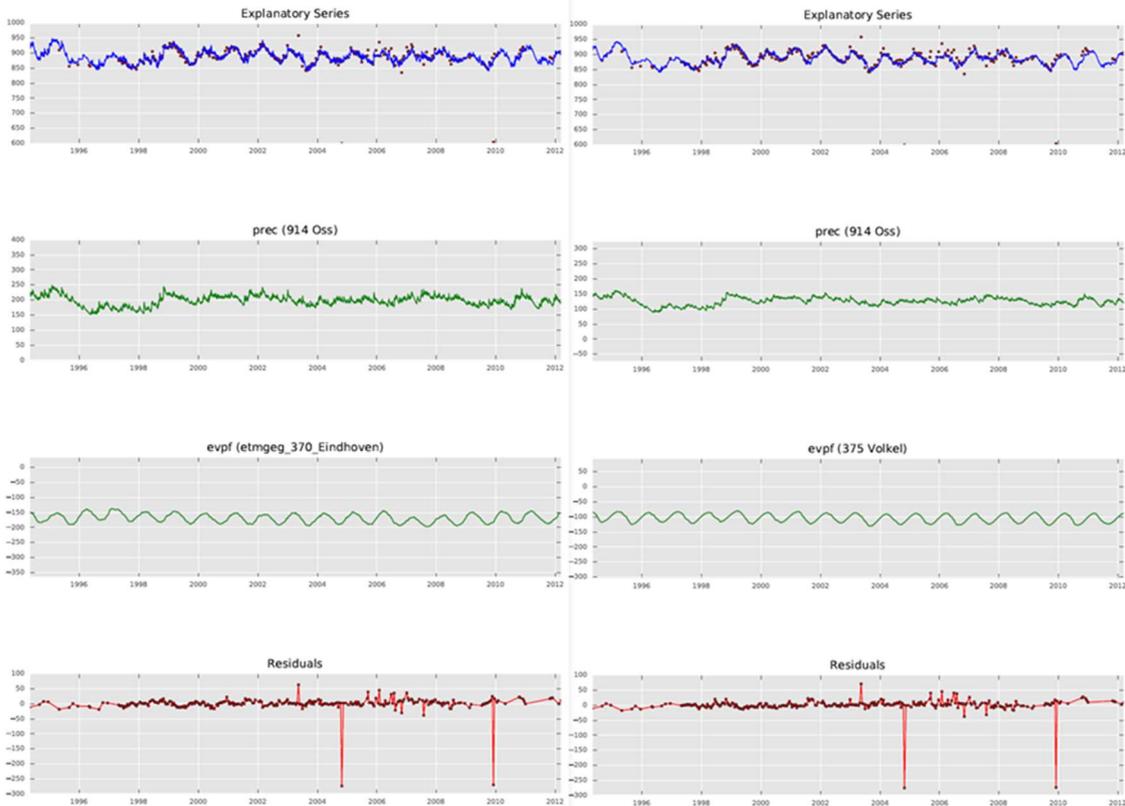


Figure 0.3 Output of model with temperature (left) and with Makkink evaporation (right) for piezometer B45E419001.

The parameter  $fc$  relates the response function for evaporation or temperature to the precipitation response. Therefore, the relation between this factor for the evaporation and temperature models will allow for a translation of degrees Centigrade into mm evaporation per day. Figure 9 gives the relation for the parameter  $fc$  in the decent models. The regression line has a slope of 0.1524 and an intercept 0.01677. We need to apply this regression to determine the equivalent average evaporation from the average temperature, if we want to use the results of a model with precipitation and temperature e.g. to calculate average recharge (as average precipitation minus average evaporation multiplied by  $fc$ ) following the method of Obergfell et al. (2019).

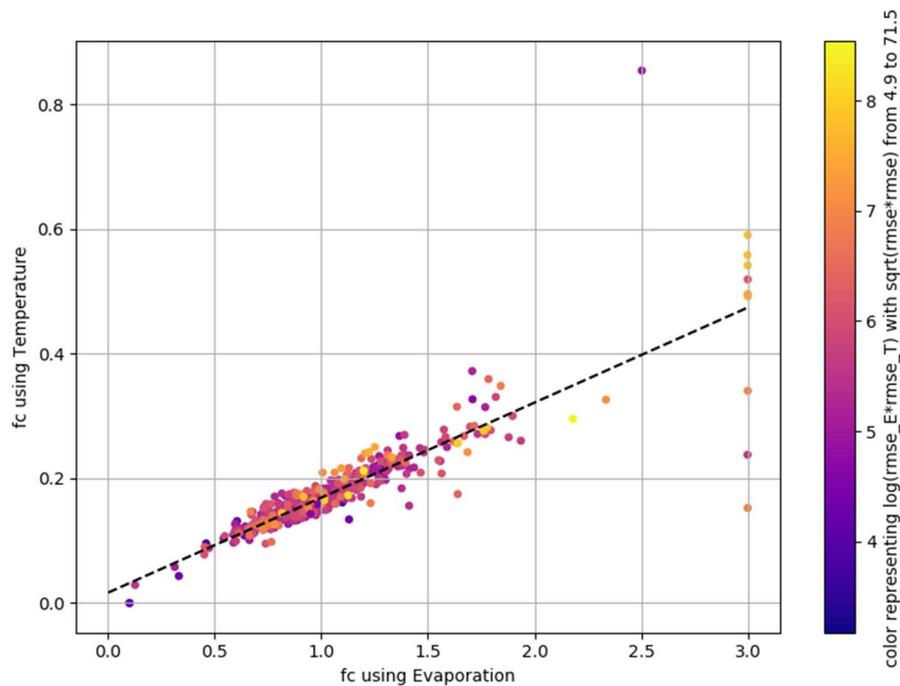


Figure 0.4 Factor ( $f_c$ ) relating evaporation or temperature response to precipitation response for decent time series models with Makkink evaporation (horizontal axis) compared to models with temperature (vertical axis) - both REGIMEOK.

Figure 0.4 shows a number of points with  $f_c$  equal to 3 for the evaporation models, which was the maximum set for this parameter. These models should be discarded because of such a physically unlikely value.

The  $f_c$  values for piezometer B51E0304001 are also far removed from the regression line: 2.503 and 0.854 for the evaporation and temperature model respectively. The time series of this piezometer does not have enough measurements to properly constrain the responses as can be seen in Figure 0.5, and this model should also be discarded.

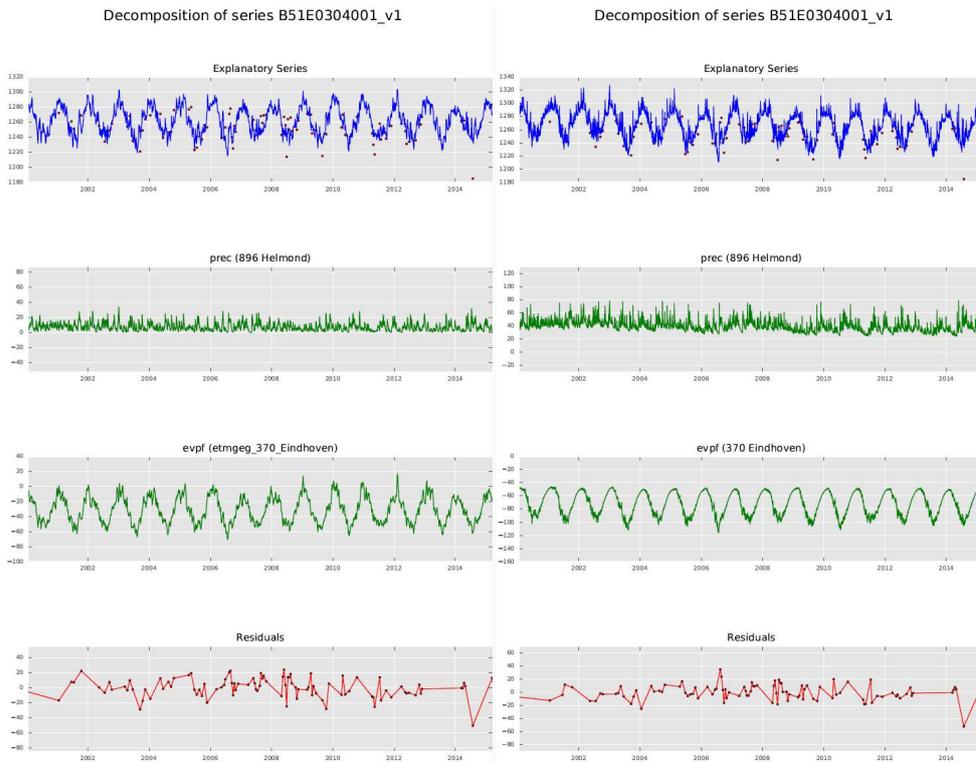


Figure 0.5 Output of model with temperature (left) and with Makkink evaporation (right) for piezometer B51E0304001.

Figure 0.6 shows that the base elevations in the evaporation and corresponding temperature models are very similar. The slope of the regression line is 1.001 and the intercept is 9.95. This shows that the conversion of temperature to evaporation derived from the regression of the fc factors is sufficient to compare temperature models with evaporation models and the temperatures with respect to zero degrees Centigrade do not give an offset in the model elevation.

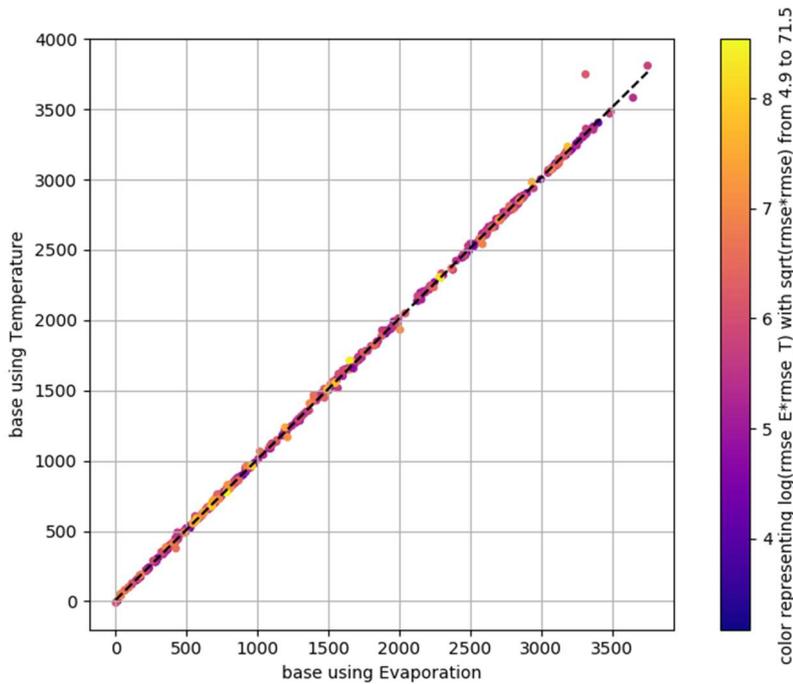


Figure 0.6 Base of decent time series models with Makkink evaporation (horizontal axis) compared to models with temperature (vertical axis) - both REGIMEOK.

Figure 0.6 has three points in the upper right corner that deviate from the regression line. These models are given in Table 0.2. The model furthest from the regression line has an evaporation factor with a value of 3 for the evaporation model, which should have been discarded as discussed before (B58D2373001). The cause of the deviation for the other two is not clear (B57A0067001 and B57A0062002).

Table 0.2 Models with base elevations greater than 3500 cm in the temperature models.

Name	E_base	T_base	E_fc	T_fc	E_M0prec	T_M0prec	E_mu_pr	T_mu_pr
B57A0067001	3648.52	3583.05	1.41	0.16	100232.7	95833.37	412.44	254.35
B57A0062002	3751.18	3810.79	0.75	0.15	92878.2	52931.16	235.33	140.04
B58D2373001	3313.47	3749.13	3.00	0.52	400978.7	461373.4	1998.39	2177.20

### Conclusions

For the analysis of the result of Metran time series models, it is necessary to only consider decent models and to discard models of lower quality. Additionally, it is good to discard models with precipitation and evaporation as input for which the evaporation factor is equal to the maximum limit value of 3 which was used in the optimization.



The precipitation response generally is not influenced by the use of temperature instead of Makkink evaporation in transfer noise modelling of groundwater head time series within the H3Oplus area in the Netherlands.

For the local climate the regression between temperature (in °C) and evaporation (in 0.1 mm/d) has a slope of 0.1524 and an intercept 0.01677, which can be used e.g. to interpret evaporation losses for such time series models where no evaporation data are available.



## B.2 Parameters of models for piezometers mentioned in text

name	models with Makkink evaporation							models with temperature						
	E_modok	E_regimeok	E_RMSE	E_M0prec	E_mu_pr	E_fc	E_base	T_modok	T_regimeok	T_RMSE	T_M0prec	T_mu_pr	T_fc	T_base
B60B0220001	1	1	8.5	12038	51	1.374	3218	1	1	9.4	10397	24	0.233	3220
B45E0419001	1	1	25.2	59752	363	1.090	865	0	1	25.0	94179	550	0.170	859
B51E0304001	0	1	11.9	18424	354	2.503	1298	1	1	12.7	3373	7	0.854	1285
B57A0067001	1	1	14.5	100233	412	1.412	3649	1	1	15.7	95833	254	0.156	3583
B57A0062002	1	1	17.2	92878	235	0.752	3751	1	1	18.9	52931	140	0.148	3811
B58D2373001	0	1	23.7	400979	1998	3.000	3313	0	1	20.6	461373	2177	0.519	3749

name	location			
	x	y	surface	topFilt
B60B0220001	191923	342113	31.8	30.96
B45E0419001	166620	418020	9.76	5.7
B51E0304001	165133	388200	14.03	13.11
B57A0067001	148106	370333	33.74	30.74
B57A0062002	148267	364624	41.28	28.16
B58D2373001	190425	356630	24.49	23.77

## B.3 Graphs of temperature and evaporation model comparison with labels

The following two graphs show comparison of temperature and evaporation models (Figure 0.7 and Figure 0.8). These are the same as in Figure 0.4 and Figure 0.6 but with the labels of the piezometers mentioned in the text.

Furthermore, the comparison of the total response and the average response time are shown (Figure 0.9 and Figure 0.10).

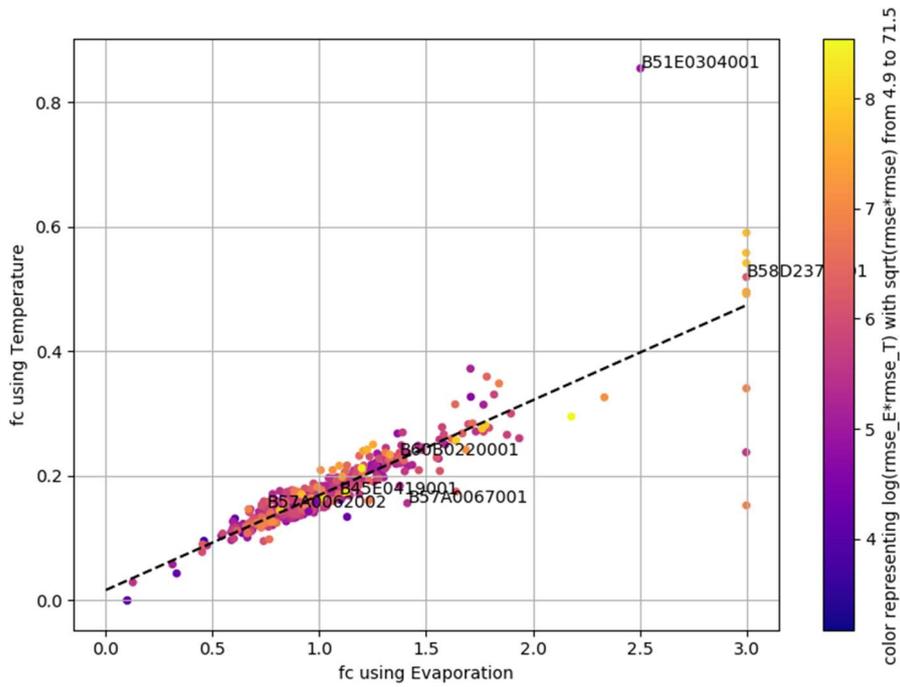


Figure 0.7 Factor (fc) relating evaporation or temperature response to precipitation response for decent time series models with Makkink evaporation (horizontal axis) compared to models with temperature (vertical axis) - both REGIMEOK.

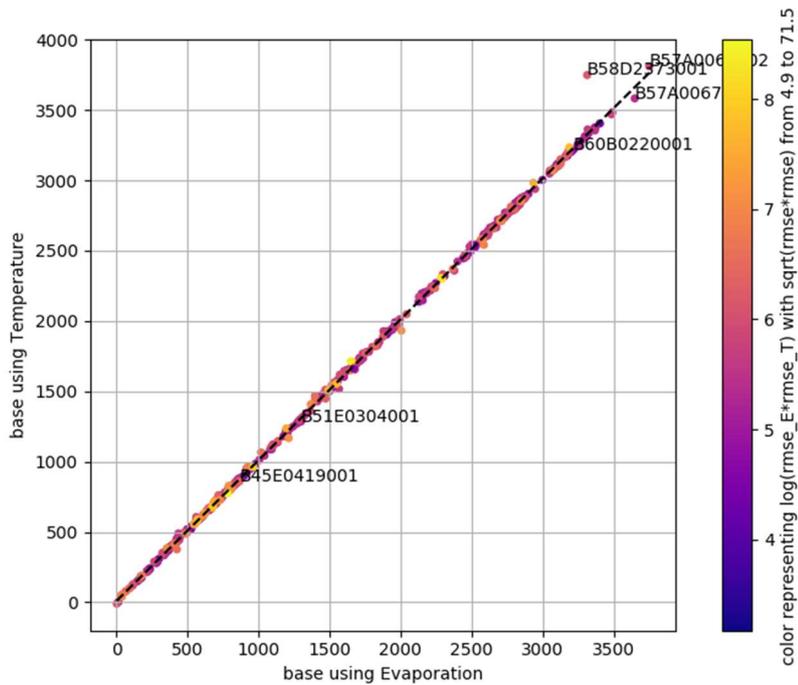


Figure 0.8 Base elevation for decent time series models with Makkink evaporation (horizontal axis) compared to models with temperature (vertical axis) - both REGIMEOK.

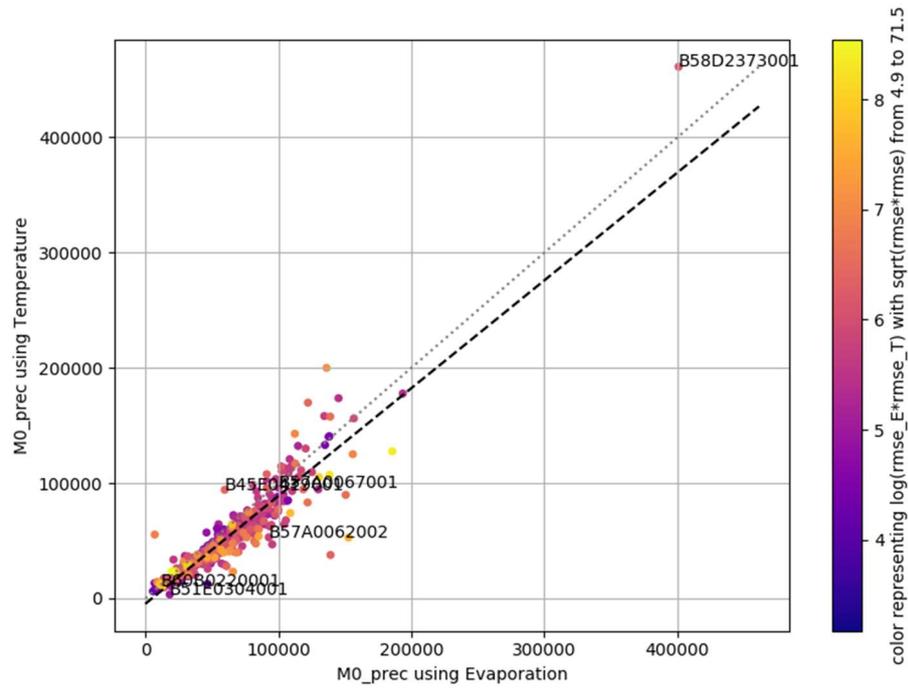


Figure 0.9 Total precipitation response for decent time series models with Makkink evaporation (horizontal axis) compared to models with temperature (vertical axis) - both REGIMEOK.

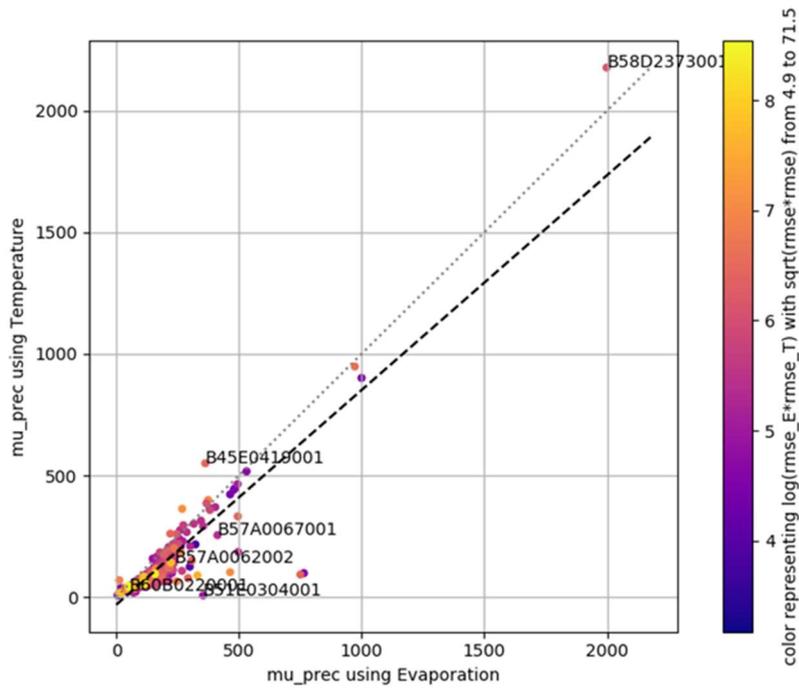


Figure 0.10 Average response time of the precipitation response for decent time series models with Makkink evaporation (horizontal axis) compared to models with temperature (vertical axis) - both REGIMEOK.

## APPENDIX C – IMPULSE-RESPONSE FUNCTION MODEL ARCHITECTURES

### C.1 Model architectures

The model architectures are identified by four letter abbreviations:

1. Input: R = single recharge; P = recharge from three principle components; M = precipitation and evaporation;
2. L indicating linear estimation of the impulse;
3. Time behaviour of impulse response function: C = constant; T = variable;
4. Use of a noise model: N = none; G = Gaussian process.

#### RLCN

A first set of four models uses a single recharge time series from the process modelling database as input, to see how well such an approach would perform. For this purpose, we selected meadow on a sand soil, with a water table at 2 m depth. Here, a constant (*i.e.* time-invariant) IRF is used, and no Gaussian process is included. The lumped error (model + measurement error) variance is inferred however, leading to the architecture illustrated in Figure 0.1. As we only have lumped error values at the dates of the level observations, we use linear interpolation to complete the unexplained time series in this case.

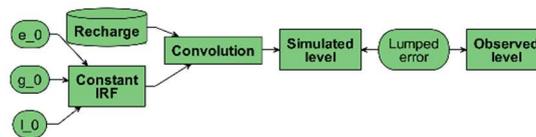


Figure 0.1 Diagram illustrating the RLCN model architecture.

#### RLCG

This variant of the RLCN architecture introduces a Gaussian process to model the unexplained part (model error + anthropogenic effects) of the time series explicitly. The parameters required for the Gaussian process are a lengthscale and a variance. The squared-exponential kernel was used here, and regularizing priors were defined for the lengthscale and variance. For the definition of the Gaussian likelihood, a fixed measurement error of 2 cm (standard deviation) was considered here. The resulting architecture is displayed in Figure 0.2.

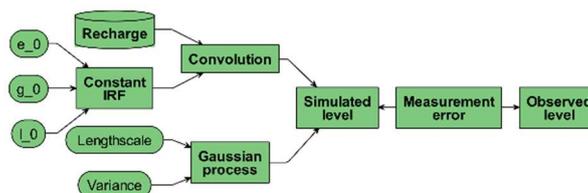


Figure 0.2 Diagram illustrating the RLCG model architecture.

### RLTN

This variant of the RLCN architecture introduces time-variance of the IRF, which means twelve parameters are added for each of the three IRF parameters, to allow for monthly fluctuations. As this does introduce a large number of parameters, the priors for these are strongly regularized. This results in the architecture displayed in Figure 0.3.

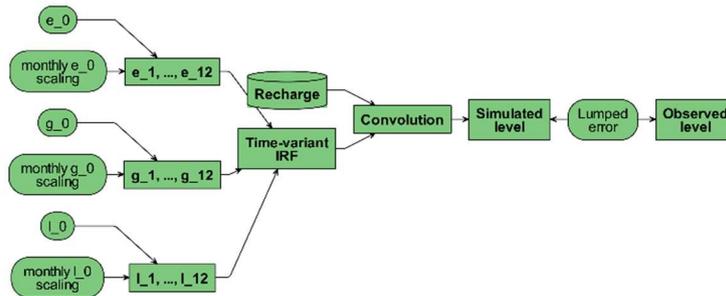


Figure 0.3 Diagram illustrating the RLTN model architecture.

### RLTG

This variant of the RLCN architecture combines both the additions in the RLCG and RLTN variants, making the IRF time-variant, and including the Gaussian process model. The resulting architecture is displayed in Figure 0.4.

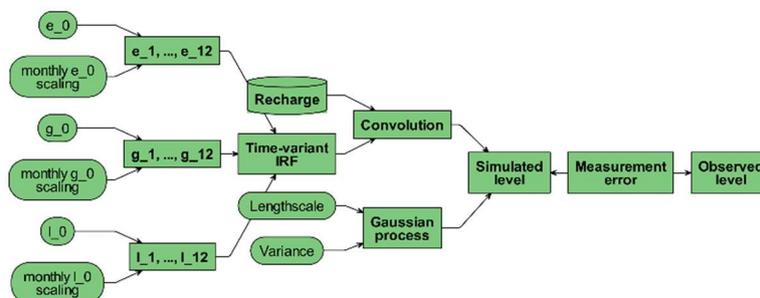


Figure 0.4 Diagram illustrating the RLTG model architecture.

### PLCN

The next set of four models starts from the complete recharge process model database, as selecting only a single time series out of the database leaves zero degrees of freedom to fit a specific groundwater level observation time series. The number of time series in the database is however 357, which is too large to work with in a transparent way. Moreover, many of the time series are heavily correlated, when there are only subtle differences in land cover, soil type and/or groundwater table depth. To overcome this, we resorted to the first three principal components of the full dataset, which explained together more than 85% of the variance in the recharge data. We then use a linear combination of these three principal components, with prior means for the coefficients corresponding to the loadings of the meadow on sand with a groundwater table at 2 m

depth case, and additional scaling to infer the mean recharge and its standard deviation as well.

While the prior mode of this approach leads to an approximation of the recharge time series used in the RLCN architecture, we do have more degrees of freedom here, which allows the model to deviate from it, and basically pick a more relevant recharge time series, somehow reflecting the conditions (land cover, soil type, and average groundwater table depth) for the groundwater level time series under investigation more accurately. In theory, we could take this even further, and make the obtained recharge depend on a first transient groundwater level estimate, but this more complex approach was considered out of scope for the current exercise (although probably useful for very shallow groundwater table depths). The resulting architecture is displayed in Figure 0.5.

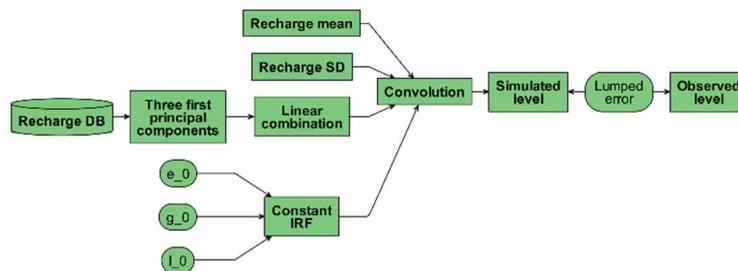


Figure 0.5 Diagram illustrating the PLCN model architecture.

### PLCG

This variant of the PLCN model again add the Gaussian process model for modelling the unexplained part explicitly, resulting in the architecture displayed in Figure 0.6.

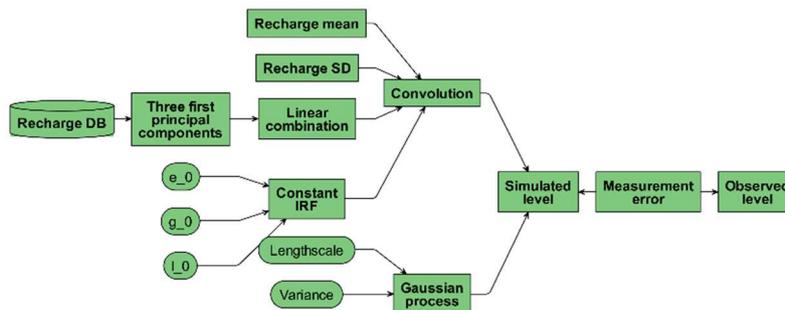


Figure 0.6 Diagram illustrating the PLCG model architecture.

### PLTN

This variant of the PLCN model again adds the time-variance of the IRF, resulting in the architecture displayed in Figure 0.7.

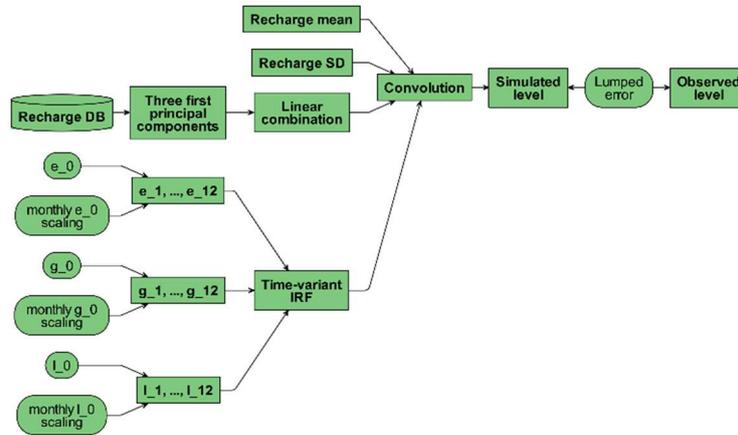


Figure 0.7 Diagram illustrating the PLTN model architecture.

### PLTG

This variant of the PLCN model combines the extra elements of both the PLCG and PLTN variants, resulting in a time-variant IRF and a Gaussian process model. The obtained architecture is displayed in Figure 0.8.

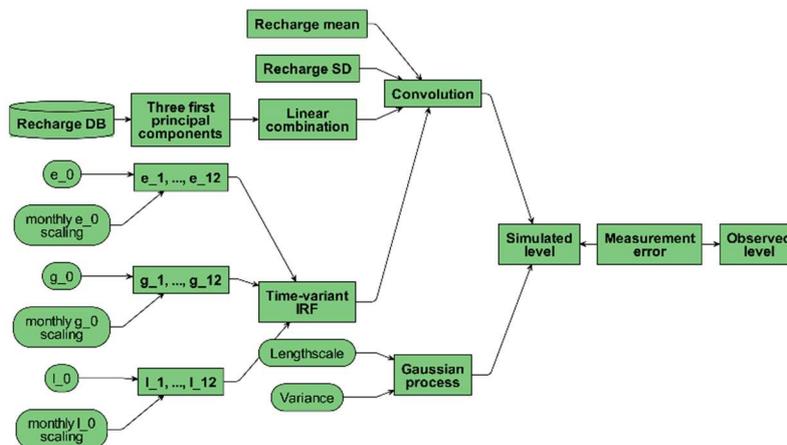


Figure 0.8 Diagram illustrating the PLTG model architecture.

### MLCN

The final set of four models uses the more traditional model inputs of precipitation and evapotranspiration. The same IRF is used for both time series, which basically means we are working with a linear combination of the two. The evapotranspiration factor is considered constant in time in this case. The resulting architecture is shown in Figure 0.9.

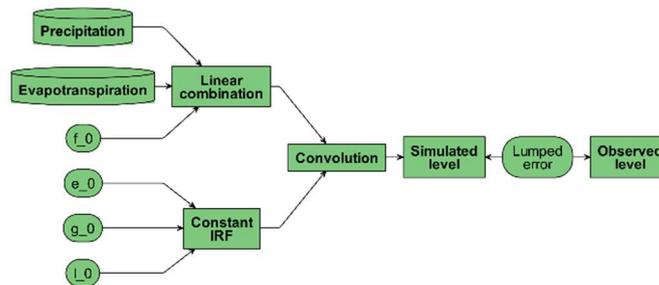


Figure 0.9 Diagram illustrating the MLCN model architecture.

### MLCG

This variant of the MLCN model add the Gaussian process to the model, resulting in the architecture displayed in Figure 0.10. This architecture is the one closest to that used by Zaadnoordijk *et al.* (2019), but still there are important differences, mainly with respect to the handling of temporal correlation in the residuals.

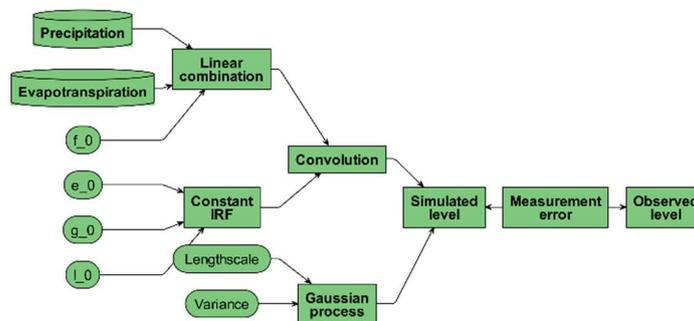


Figure 0.10 Diagram illustrating the MLCG model architecture.

### MLTN

This variant of the MLCN model add the time-variance of the IRF. We consider in this case also the evapotranspiration factor to be time-variant, as suggested by Collenteur *et al.* (2021), resulting in again twelve extra, but strongly regularized, parameters. The resulting architecture is displayed in Figure 0.11.

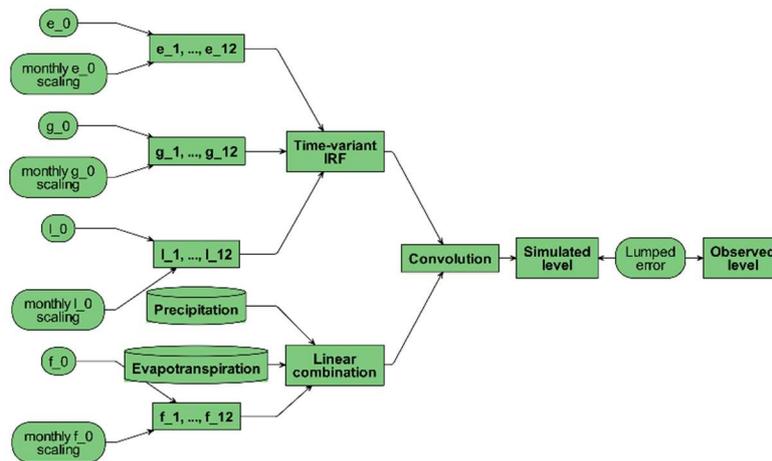


Figure 0.11 Diagram illustrating the MLTN model architecture.

### MLTG

This variant of the MLCN model again combines the additions of both MLCG and MLTN variants, resulting in time-variance of the IRF and evapotranspiration factor, as well as a Gaussian process model. The resulting architecture is displayed in Figure 0.12.

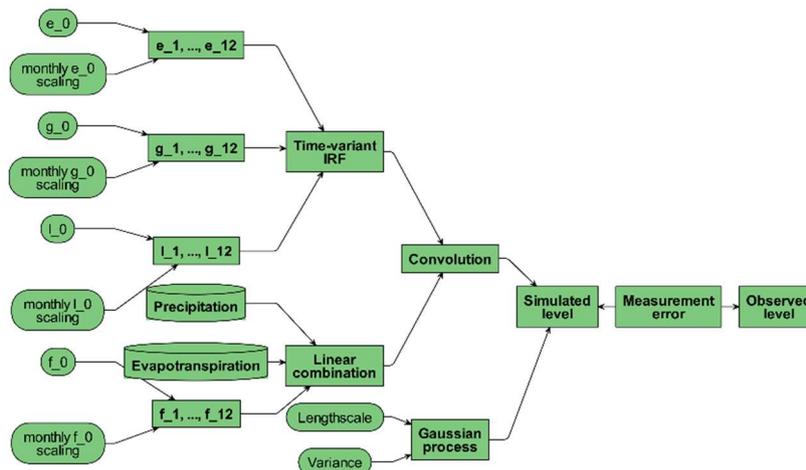


Figure 0.12 Diagram illustrating the MLTG model architecture.