University of Tartu
Faculty of Science and Technology
Institute of Ecology and Earth Sciences
Department of Geography

Master's thesis in Geoinformatics for Urbanised Society (30 ECTS)

Evaluating impulse response function models for groundwater level modeling across the Baltic states

Marta Jemeļjanova

Supervisors:
PhD Alexander Kmoch

MSc Jānis Bikše
Faculty of Geography and Earth Sciences, University of Latvia
MSc Raoul Collenteur
Institute of Earth Sciences, University of Graz (Austria)

Tartu 2022

Abstrakt

Põhjaveetaseme modelleerimine sademetepõhiste mudelite abil Balti riikides

Põhjavee põudade modelleerimiseks ja iseloomustamiseks on vaja korrigeeritud aegridu pidevate ajaliste sammudega. Pikaajalised põhjaveetaseme andmestikud sisaldavad aga nii vigu kui ka aastatepikkusi lünki. Põhjavee taseme sademetele reageerimise funktsioone (TFN-IRF) kasutavad aegridade mudelid on tänu madalale parameetrite arvule ja kiiretele arvutustele väärtuslik vahend päevaste põhjaveetasemete aegridade saamiseks. Antud töö eesmärkideks oli määrata aegridade mudelite sobivust impulssreaktsiooni funktsioonidega ning analüüsida erinevate keskkonnakirjelduste seoseid mudeli sobivuse ja mudeli parametriseerimisega.

Mudeli sobivust hinnati kombineerides erinevaid mõõdikuid ja visuaalse hindamise teel. Seoseid analüüsiti korrelatsiooni- ja lineaarse regressioonianalüüsi abil. Mudeli täpsus varieerus mõõtmisalade vahel, kuid oli kõrgem kaevudes, kus kogu modelleeritud aja jooksul oli veetaseme muster ühtlasem. Parima tulemusega mudelid saadi piirkondadele, mis asusid topograafiliselt kõrgemal, kus aasta keskmine sademete hulk oli suurem. Mudeli parameetrite väärtused korreleerusid nii mõõtmisalade topograafia kui geoloogiaga.

Märksõnad: põhjavee modelleerimine, impulssreaktsiooni funktsioonid, füüsilis-geograafilised omadused.

CERCS kood: P470 – Hüdrogeoloogia, geoplaneering ja ehitusgeoloogia

Abstract

Evaluating impulse response function models for groundwater level modeling across the Baltic states

For modeling and characterizing groundwater droughts, corrected time series with constant time steps are required. However, long-term groundwater level datasets contain both errors and years-long gaps. Time series models using impulse response functions (TFN-IRF) are a valuable tool in obtaining such daily groundwater level time series due to low parameter count and quick calculation time. The objectives of the work were to determine the obtainable fit of time series models with impulse response functions and analyze the relationships between various environmental descriptors and model fit and model parametrization, respectively. The model fit was evaluated by the combination of various metrics calculation and visual evaluation. The relationships were analyzed by correlation and linear regression analysis. The achievable model fit varied site by site but was higher in wells with uniform patterns throughout the modeled time frame. While a higher achievable model fit is characteristic of locations higher on the local topography and with higher average yearly precipitation, the model parameter values correlated with both site topography and geology.

Keywords: groundwater modeling, impulse response functions, physiographic descriptors.

CERCS code: P470 - Hydrogeology, geographical and geological engineering.

Table of contents

1. Introduction

Groundwater has a significant role in ecosystem services sustaining dependent ecosystems as well as storing water long term (Griebler, Avramov 2015). Additionally, it is a major source of drinking water in various countries (European Environmental Agency 2018). The groundwater level is determined and affected by various hydrogeological properties together with meteorological conditions as well as anthropogenic factors (water abstraction being the biggest reason for local drawdown). Current predictions for the future foresee increased anthropogenic groundwater withdrawal due to higher demand, decreased recharge of groundwater, and negative impact on groundwater-dependent ecosystems (Wada et al. 2010; European Environmental Agency 2018), but the extent of it is still unknown and being studied.

Many types of analyses aim to contribute to this research field, including the study of future groundwater droughts. For modeling and characterizing the droughts, homogenous and corrected time series are necessary. Long-term groundwater level datasets, however, contain various types of limitations, for example, the measurement frequency is varied both spatially and within the same well (from a couple of times per year to bi-daily). The measurements in wells span different periods and have years-long gaps. In addition, the groundwater level must be predicted for the future period of the study.

Time series models using impulse response functions (TFN-IRF) are a valuable tool in obtaining such daily groundwater level time series due to low parameter count, quick calculation time, and relatively high model fit as compared to other groundwater level models (von Asmuth et al. 2008). The TFN-IRF modeling has not yet been applied to a high count of groundwater level time series in the Baltic states. Furthermore, the relationships between the environment of the observation wells and model fit and model parameters have not yet been extensively studied.

Aim: to model groundwater levels of observation wells in the Baltic states using TFN-IRF modeling and to evaluate the suitability of the method.

Research questions:

1. What is the fit that can be achieved using time-series modeling with impulse response functions on groundwater level and meteorological data of the Baltic states?
2. Which site characteristics of the observation wells impact the fit of the time-series models with impulse response functions?
3. Which site characteristics of the observation wells impact the parameters of the time-series models with impulse response functions?

This thesis is comprised of four chapters. The first chapter consists of the groundwater level determining factor description and an overview of commonly used hydrological models. The second chapter details the data and methods used in this thesis. The results are presented in the third chapter.

In the fourth chapter, the obtained results are discussed, and the applied methods evaluated. The conclusions drawn from the study are written at the end of the fourth chapter.

2. Theoretical overview

2.1. Groundwater and its affecting factors

Groundwater comprises 95% of unfrozen freshwater sources (Alley et al. 2006) making it the most important freshwater source. The amount of groundwater resources is determined by the difference between rainfall (inflow) and runoff and evaporation (outflow). In addition, it is fed by both surface and subsurface flows (Freeze, Cherry 1979). The volume of groundwater in a specific location consequently fluctuates based on meteorological and climatic events (Fitts 2002). The residence time of groundwater can be of various lengths – from days to tens of thousands of years (Alley et al. 2006).

Various tests are performed to determine groundwater flow characteristics. Tracer (injection) tests measure the tracer compound concentration in monitoring locations based on time. From these time series, called breakthrough curves, the aquifer transport parameters are calculated (Ptak et al. 2004). These tests contain various assumptions and simplifications of the system, and the calculation is further complicated by the background values and continual input from rainfall of some of the used compounds (e.g., ^{14}C, ^{3}He) (Cartwright et al. 2017). Similarly, pumping tests make use of time series curves for parameter estimation. The results of such tests are significantly affected by the chosen methodology of calculation and the distance between a monitoring location and the pumping location (Hantush 1956).

The groundwater level also called the head or hydraulic head is the level of the intersection between water and atmosphere in the observation well using a specified datum (Post et al. 2007; Fitts 2002). The groundwater level is measured using wells (see Figure 1) or piezometers (that have a smaller diameter than wells) and sometimes are coupled with pressure transducers. The upper end of the well is open to the atmosphere while the groundwater flow is fostered through holes in the lower part of the well (Fitts 2002). In recent decades, manual measuring has been taken over by automatic data logging (e.g., Retiķe et al. 2022).

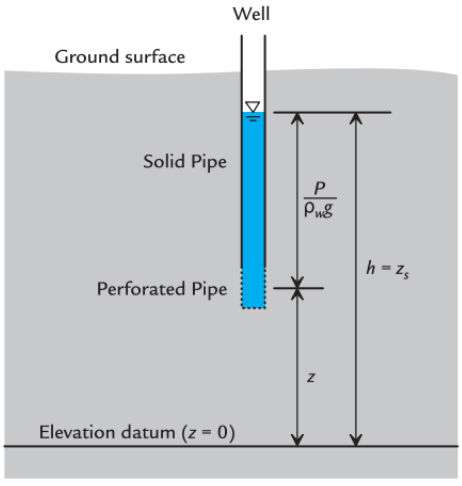

Figure 1. Groundwater level measuring in an observation well (Fitts 2002).

The factors impacting groundwater head variations can be divided into five principal groups – geological and hydrogeological, geomorphological, hydrometeorological, anthropogenic, and cosmogenic (solar and geomagnetic activity) (Zektser, Loaiciga 1993). Factors can impact the groundwater for various time frames, and some examples can be seen in Table 1.

Table 1. Factors and their impact time on the groundwater level (adapted from Freeze, Cherry 1979).

Factor	Type of aquifer impacted	Type of source	Length of impact
Groundwater recharge	Unconfined	Natural	Seasonal
Evapotranspiration	Unconfined	Natural	Short-lived
Tidal effects in the vicinity of waterbodies	Unconfined	Natural	Diurnal
Groundwater pumping	Unconfined and confined	Anthropogenic	Long-term
Agricultural irrigation	Unconfined	Anthropogenic	Long-term

Hydrometeorology

Climate variables, such as precipitation, temperature, and evaporation have a significant impact on groundwater level fluctuations since groundwater level tends to mimic the respective Köppen-Geiger climate class's intra-annual variability (Barron et al. 2012; Nygren et al. 2020). For example, for the cold climate type the groundwater level fluctuations are driven by winter snowfall, spring snowmelt, intermittent rainfall and high evapotranspiration in summer, and rains in autumn (Nygren et al. 2020). Where snow cover would result in groundwater level decline due to reduced recharge, a spring flood is responsible for a significant increase in recharge (Nygren et al. 2020; Van Loon et al. 2010).

The meteorological drought propagates to groundwater in the form of decreased recharge caused by precipitation deficit among other reasons (Zaidman et al. 2002; Van Loon 2015). The persistence of groundwater drought varies based on precipitation patterns, aquifer transmissivity, and effective porosity lasting from mere days to years (Apurv et al. 2017; Soleimani et al. 2017; Van Loon 2015), but the onset can be lagged as compared to the meteorological drought (Van Loon 2015; Babre et al. 2022). The impact of drought on groundwater tends to be significantly longer than on runoff, but the count of separate drought instances – lower (Sutanto, Van Lanen 2021). The most significant drought events in the Baltics between the years 1989 and 2018 took place in the years 1992-1994, 1996-1997, 2002-2004, 2005-2007, and 2018 with a one-month delay from the meteorological drought instance (Babre et al. 2022). Future predictions indicate lower recharge of groundwater due to both decreased precipitation and increased evaporation. With decreasing snow cover, the peak recharge could happen

earlier in the year (Luoma, Okkonen 2014) and spring snowmelt would exceedingly be substituted by winter rainfall (Nygren et al. 2020).

Geology and hydrogeology

Water saturated geologic strata are divided into three groups based on the capacity of groundwater transmissivity and permeability – aquifers (high transmissivity and permeability), aquitards (lower permeability), and aquicludes (no transmissivity). In describing a local aquifer system, the geologic strata can be distinguished comparatively, by strict criteria (Freeze, Cherry 1979) or by referencing the aquifer yield (Lerner, Harris 2009).

The porosity and primary and secondary permeability depend on the lithological properties (the physical characteristics) of the strata. Groundwater in more permeable strata (such as sand) responds quicker to the precipitation and vice versa (Wendland et al. 2008; Lorenzo-Lacruz et al. 2017; Allen et al. 2010). Strata consisting of carbonate rocks, although having low permeability, are prone to faults and fractures. Faults (corresponding to the structure of the strata) in such cases can either increase the groundwater flow to an extent that the strata are considered an aquifer or act as a local barrier (Bense et al. 2013). Therefore, where aquifers generally consist of terrigenous sediments (gravels, sands), fractured rocks, and limestones, aquitards consist of clays and shales. Lithology can have both local and regional differences, for example, Scandinavia mostly consists of low permeable, low porosity aquifers and non-aquiferous rocks, while the central latitudes of Europe (Northern and central parts of Germany, Poland) have moderate to high permeability unconsolidated sediments of high porosity (Cornu et al. 2013; Duscher et al. 2015).

The response to the precipitation, however, is additionally impacted by the aquifer storage capacity, with higher storage capacity contributing to slower response and recession of the groundwater levels (Allen et al. 2010). Aquifers, in addition, can be divided into unconfined and confined between aquitards (Freeze, Cherry 1979), with either having different characteristics and therefore also response to the impulse of precipitation (Fitts 2002, Rasmussen, Crawford 1997). Stratigraphy which describes the age and geometrical aspects (depth, thickness) of the formations is the third determinant of groundwater flow in geological strata (Freeze, Cherry 1979).

Topography

Groundwater depth, flow, and recharge is strongly impacted by topography (Condon, Maxwell 2015; Hokanson et al. 2019), namely slope and relief (Forster, Smith 1988). Groundwater flows from recharge to discharge areas – mainly from highlands to valleys, where the flow is downward in recharge areas and upward in discharge areas (Freeze, Cherry 1979). Multiple flows can be discerned – local, regional, and intermediate (see Figure 2), and they are determined by local relief – in the territories of significantly varied relief, local systems develop, otherwise regional flows are developed (Freeze, Cherry 1979). Topography, together with soil permeability, is the main predictor of time to groundwater rise as well as recession. Considering locations with low permeability soils, the response

time to the precipitation input is faster for locations with lower slope of the water table and higher upstream contributing area, but the opposite for locations with highly permeable soils (Rinderer et al. 2016).

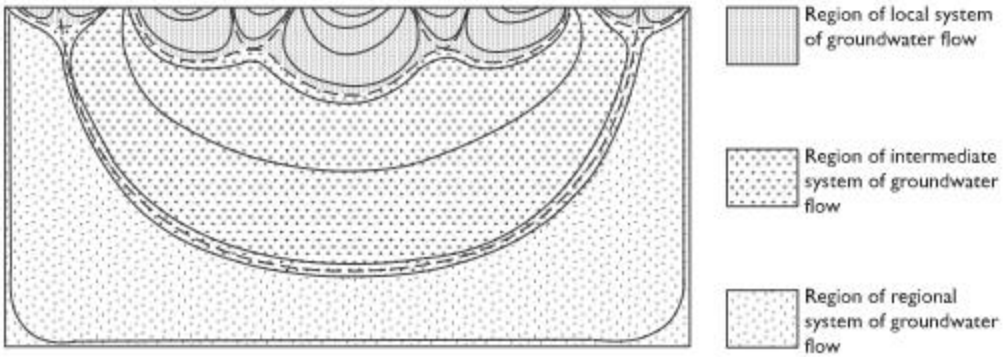

Figure 2. Local, regional, and intermediate systems of flow (Freeze, Cherry 1979).

Surface waterbodies can act as recharge and discharge areas for regional groundwater depending on topography (Alley et al. 2006; Freeze, Cherry 1979). Streams, lakes, and wetlands depending on the head gradient with groundwater and regional flow characteristics can be either gaining recharge, discharging, or mixed pattern. In the case of severe weather instances however, such as heavy rainfall, the relationship can change (Kaleris 1998; Alley et al. 2006). Hence why groundwater level in the vicinity of e.g., streams, exhibits similar fluctuation patterns of the waterbody (Liu et al. 2007).

Land use change and anthropogenic impact

Various anthropogenic activities impact the local groundwater level (Kløve et al. 2011). For example, the groundwater level decreases when a natural grassland or forest is cultivated (Kroes et al. 2019; Zhang, Hiscock 2010) and decreases when the territory is deforested (Hotta et al. 2010; Smerdon et al. 2009). A substantial impact is connected to withdrawing water for, e.g., irrigation purposes (Nistor 2020). Excessive abstraction can not only decrease the quality and quantity of groundwater resources but also cause degradation of groundwater-dependent ecosystems (e.g., wetlands) (Lerner, Harris 2009). Around the coastal areas, groundwater abstraction causes saltwater intrusion if the hydraulic head is low enough, therefore decreasing the quality of the groundwater (Raidla et al 2019). The predicted increase of sea water level due to climate change could further exacerbate this problem (Luoma, Okkonen 2014; Eriksson et al. 2018) while the overall degree of abstraction is expected to increase (Amanambu et al. 2020).

All three Baltic states have seen an increase in groundwater levels after the collapse of the Soviet Union (Karro et al. 2009; Levins et al. 1998; Juodkazis 1994). The depression funnels began developing in the 1950s due to a significant increase in water abstraction as a response to the increasing industry and expanding centralized water supply among other reasons (Klints, Dēliņa 2012). Another outcome

caused by excessive groundwater withdrawals was (and currently still is) sea water intrusions in the vicinity of the seashore, namely Liepāja (Levins et al. 1998; Bikše, Retiķe 2018), Kolka and Carnikava in Latvia (Dēliņa 2018) as well as Viimsi island in Estonia (Raidla et al. 2019). Draining peatlands was also characteristic across the Baltic countries during Soviet times (Karofeld et al. 2017), which results in changed hydrological characteristics – faster response to the impulse of precipitation and lower water storage capacity (Ahmad et al. 2020). Various restoration activities have been undertaken since (Karofeld et al. 2017; Pakalne et al. 2021; Küttim et al. 2018).

2.2. Hydrological modeling

Hydrological models are simplified representations of hydrological systems (Solomantine, Wagner 2011) and are used to learn more about the behavior of hydrological systems, assess potential future scenarios or fill data gaps (Silberstein 2006). The models are classified by their characteristics – based on their structure, the modeled timeframe, the extent of space, and model stochasticity (involved randomness) (see Figure 3).

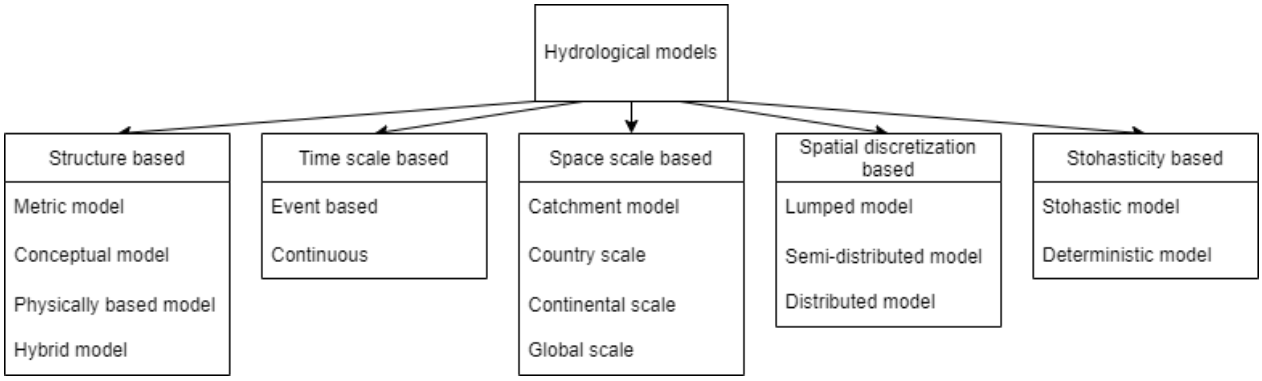

Figure 3. Classification scheme of hydrological models (adapted from Paul et al. 2021).

Model structure defines the level of reliance on actual physical processes in the calculations. Metric models, also called empirical, black box, and data-driven models (Solomatine, Wagner 2011), are sets of equations describing the relationship between observed and modeled variables ignoring the real-world processes and relationships in the catchment, hence they are rather hard to interpret and are site-specific (Devia et al. 2015). In case of significant system changes, metric models might offer significantly weaker, if not entirely incorrect, results (Solomatine, Wagner 2011). Conceptual models, also called gray box and parametric models, contain physically based equations, but still contain a calibration step for certain parameters. For example, in a model based on a system of reservoirs (filled with rainfall and emptied by evaporation) the reservoir size needs to be calibrated. Physically-based models, also called white-box models, are more complex than conceptual models and need more input data, but still can contain parameters to be calibrated. These models are suggested in case of high spatial detailing (Devia et al. 2015; Solomatine, Wagner 2011).

The models can be applied to varying spatial scales. While lumped models consider a point in space (Becker, Serban 1990), semi-distributed models implement physical processes and consider a catchment as multiple sub-catchments. They require less data than distributed models that consider the catchment as a whole (Jajarmizadeh et al. 2012; Solomatine, Wagner 2011).

Hydrological and meteorological processes are a sum of both periodic and stochastic effects, for example, long-term precipitation series have an evident periodic mean, but also an independent stochastic element (Yevievich 2012). A deterministic model describes all parameters with mathematical equations following cause-effect relationships. Stochastic models, however, use probability equations since the relationships do not adhere to simple cause-effect relations (Jajarmizadeh et al. 2012; Yevjevich 1987). Results consequently include a level of randomness (Solomatine, Wagner 2011). An example of a stochastic model is the Autoregressive moving average (ARMA) model (Yevievich 2012) which is a composition of both past observations and errors (Makridakis, Hibon 1997).

The choice of model depends on the specifics of the study, as the model complexity does not always correlate with better results. In addition, more complex models can suffer from overparameterization (Orth et al. 2015). Beven (1993) emphasizes that the more complex distributed models are rather a "prophecy" than a prediction because they also suffer from certain drawbacks. At the same time, lumped models require significantly less data and less calibration (Bakker, Shaars 2019), however, they can underrepresent the relationships. Ultimately, the choice depends on the modeled event, e.g., different models depending on if flood or drought is modeled (Orth et al. 2015) or a more complex model in case of studying an ungauged site as compared to sites with monitoring data and a unified response, where a lumped model could be appropriate (Refsgaard, Knutsen 1996). For modeling, various types of input data are used, such as geological, hydrological, soil, and meteorological data (Singh 2018), to name a few, from various origins, for example, ground measurements (e.g., Marchant, Bloomfield 2018) or reanalyzes. The models can cover either a single event or multiple years from the spatial scale of a catchment to global (Jajamizadeh et al. 2012).

Models used for groundwater level reconstruction and prediction

The Soil and Water Assessment Tool (SWAT) is a semi-distributed process-based model meant for river basin scale with the initial aim of aiding the understanding of land management impact on waters and soil (Arnold et al. 1998). The model was first developed by the USDA Agricultural Research Service (Gassman et al. 2007). SWAT generally is a versatile model with large-scale applications over different territories (Gassman et al. 2007): SWAT's applicability includes, but is not limited to, determining daily streamflow (Zeiger et al. 2021) and runoff (Ahmadzadeh et al. 2022), nutrient flows (Yuan, Koropeckyj-Cox, 2022), subsurface drainage (Bailey et al. 2022), and groundwater table (Melaku, Wang 2019), considering a wide scale of factors such as varying climate (Ahmadzadeh et al. 2022). The recognized drawbacks of the model are as follows: decreased accuracy when using time-step less than daily, the model is adjusted to the U.S. environment, and the model is not automatically

updated with new incoming data (Wang et al. 2019). Although SWAT considers groundwater, it is not thoroughly developed for such a purpose (Kumar et al. 2021).

MODFLOW is a modular hydrologic simulation program (Langevin et al. 2017) originally developed by the US Geological Survey in the 1980s as a groundwater flow modeling program (McDonald et al. 2003). Currently, its capabilities have been extended to modeling surface water flow and solute transport. It can be enhanced with various modules, for example, Groundwater Flow Model (Langevin et al. 2017). The newest core version - MODFLOW 6 - is written in Fortran programming language. The principal scheme of the program consists of four basic components: the model solves a hydrologic process (or multiple different processes), and a solution solves one or multiple models iteratively. An exchange facilitates the hydraulic connection between models, and the timing module is responsible for time steps and cutoff time for simulation (Hughes et al. 2017). There are numerous use cases for MODFLOW in agriculture, drought studies, sea water intrusion, and others (Hariharan, Uma Shankar 2017). The defined problems modeling groundwater specifically are the underestimation of infiltration flux or considering uniform under rivers (which is not always the case). There could be mismatches between river and cell sizes that affect the infiltration rate. Both can give biased results (Brunner et al. 2010).

Aquimod is a lumped conceptual model created by the British Geological Survey. It was written in the C++ programming language. The model consists of three modules – soil zone, unsaturated zone, and saturated zone. The soil zone model divides the incoming precipitation into runoff, evaporation, drainage, and soil storage. It conceptually mimics the bucket system. The unsaturated zone is represented with a Weibull probability density function, but the groundwater head in the saturated zone is determined by mass balance calculations. Depending on the specific geologic properties of the site or the aim of modeling, any of the parts can be omitted. It consists of 13 parameters that are calibrated using a Monte Carlo approach. According to the authors, the model is capable to replicate sinusoidal patterns quite well, but the results can be worse for more complex fractured hydrological systems. In addition, the model is incapable to simulate events that it hasn't been calibrated over, such as extreme drought or wetness (Mackay et al. 2014). The example use cases of the model are short-term groundwater level forecasts (up to three months) (Mackay et al. 2015), simulating groundwater levels in unconfined aquifers (Mackay et al. 2014), and reconstructing historic groundwater levels (Jackson et al. 2016), among others.

HYDRUS model is an open-source numerical model for modeling multiple processes concerning both soils and groundwater (e.g., infiltration, groundwater recharge, contaminant transport among others) (Simunek et al. 2012). The model solves Richard's equation for water flow and advection-dispersion equations for the transport of heat and solutes (Pietrzak 2021). Richard's equation describes the water flow through porous media due to capillary and gravity forces (Farthing, Ogden 2017). Where HYDRUS-1D considers only one dimension, HYDRUS-2D/3D has the capability of 2 and 3 dimensions, thus extending the possible scale, although catchment-wise modeling is not recommended (Pietrzak 2021). The model can be extended by various modules, such as a wetland module, C-Ride module (colloid transport), and DualPerm module (Simunek et al. 2013).

Transfer function-noise models (TFN) with impulse response functions, also called time series models, translate input series, such as precipitation, into output series (groundwater level) with the help of plausible impulse response functions. The functions quantify and describe the impact of the impulse, or the input series, on the output series (Bakker, Schaars 2019). TFN models, due to the specifics of IRF, can be used to characterize certain properties of local hydrology and the response time of the groundwater level to stresses (Pezij et al. 2020). TFN-IRF models can be used for multiple purposes, such as filling gaps in groundwater level time-series (von Asmuth et al. 2008), estimating groundwater recharge (Collenteur et al. 2021) cleaning time-series of observed groundwater levels - finding outliers (erroneous measurements) (Peterson et al. 2018), detecting changes in groundwater variability regime (Obergfell et al. 2019), and forecasting groundwater levels – simulating levels in the future based on climatic predictions (Brakkee et al. 2021). In addition, they can be used to analyze the variability of groundwater levels, and separate and quantify the effect of each stress on the GWL (Bakker, Shaars 2019). Possible causes of low model fit are missing stresses that were not added to the model but have a significant impact on the level (such as large-scale abstraction), changed system behavior that the IRF cannot replicate (e.g., a linear recharge model), and input time series that are not representative enough – they are too short, or the frequency is too low (Zaadnoordijk et al. 2019).

The groundwater level is influenced by multiple environment factors and to a varying intensity and persistence. The models that are used to characterize and simulate the groundwater levels have different capabilities and limitations. In this study, the time series models with impulse response functions will be used to model groundwater levels due to their capability of gap filling and the links between environment and model fit or parametrization, respectively, will be studied.

3. Data and methodology

3.1. Study area

The three Baltic countries – Estonia, Latvia, Lithuania – are located in Northeastern Europe with a total area of 175 117 km^2 (Kriauciuviene et al. 2012). The countries belong to *Dfb* (snow climate, fully humid, warm summer) climate type according to the Köppen-Geiger classification (Kottek et al. 2006; Beck et al. 2018), but the local climate varies from maritime (western part, due to the Baltic Sea) to continental (eastern, the southeastern part of countries, inland) (Kriauciuviene et al. 2012). There are additional temperature variations along the North-South gradient as well as based on the orography i.e., decreased temperature and increased precipitation in highlands (Bethere et al. 2017), although the highest peaks are only just over 300 m. The annual precipitation varies from under 600 mm to over 800 mm (Jaagus et al. 2010) and there is a high prevalence of S-SW winds (Pogumirskis et al. 2021).

3.1.1. Hydrogeological setting

Geologically the countries belong to Baltic Artesian Basin (BAB) which is rich in groundwater supply (Juodkazis 1994). The crystalline basement of BAB consists of Paleoproterozoic rocks, sloping by 2-4 km^{-1} southwards, that across Baltic countries are covered with terrigenous and carbonate sedimentary rocks and subsequent overlay of Quaternary deposits (Kitterød et al., submitted; Raidla et al. 2009) (see Figures 4 and 5).

The **Cambrian-Vendian** aquifer system covers all of Estonia, Latvia, and Lithuania (Figures 3 and 4), except the Mõniste-Lokno uplift (Southern Estonia). It is comprised of sandstones and siltstones interlaid with clay (Perens, Valner 2007). The depth of the Cambrian-Vendian aquifer system in Northern Estonia is 60-70 m, in Latvia, it ranges between 400-1600 m (Levins et al. 1998). The Voronka-Gdov aquifer, positioned in western Estonia, is divided into separate Gdov (over the Precambrian basement) and Voronka aquifers to the east by the Kotlin formation (Vaikmäe et al. 2001; Perens, Valner 2007). Both the Kotlin formation and the Gdov aquifer extend to Latvia. While the Kotlin formation covers all of Latvia, the Gdov aquifer does not cover the southeast (Levins et al. 2012).

The aquifers are overlaid by the Lontova aquitard consisting of clays (eastern side) and the Voosi formation (western side) (Raidla et al. 2009; Perens Valner 2007). The Cambrian-Vendian aquifer system is the main drinking source in Northern Estonia (Karro et al. 2009), but due to the high mineralization and slow flow is called the “stagnant” or slow water exchange zone in Latvia and Lithuania (Kitterød et al., submitted).

Figure 4. The geological cross-section line across the Baltic countries (adapted from Krūmiņš et al. 2021).

The separately distinguished **Ordovician-Cambrian** system in Estonia covers all the territory except for the northern shore, Mõniste-Lokno uplift, and the islands. The aquifer consists of fine-grained sandstones and siltstones, and its thickness reaches up to 60 m, increasing southwards (Perens, Valner 2007).

The **Silurian-Ordovician** aquitard overlays the Cambrian-Vendian aquifer system (Kitterød et al., submitted). It is comprised of marls, limestones, siltstones, and argillites and its maximum thickness within Estonia is 350 m (Perens, Valner 2007), while in Latvia and Lithuania it reaches up to 800-850 m in thickness (Kitterød et al., submitted; Levins et al. 1998). The Silurian-Ordovician aquifer system consists of limestone and dolomite with clay interlayers. The aquifer system consists of zones with lateral and vertical water movement in fissures and the upper part is considered a significant water source for the central, western Estonia, and islands of the archipelago (Perens, Valner 2007; Karro et al. 2009).

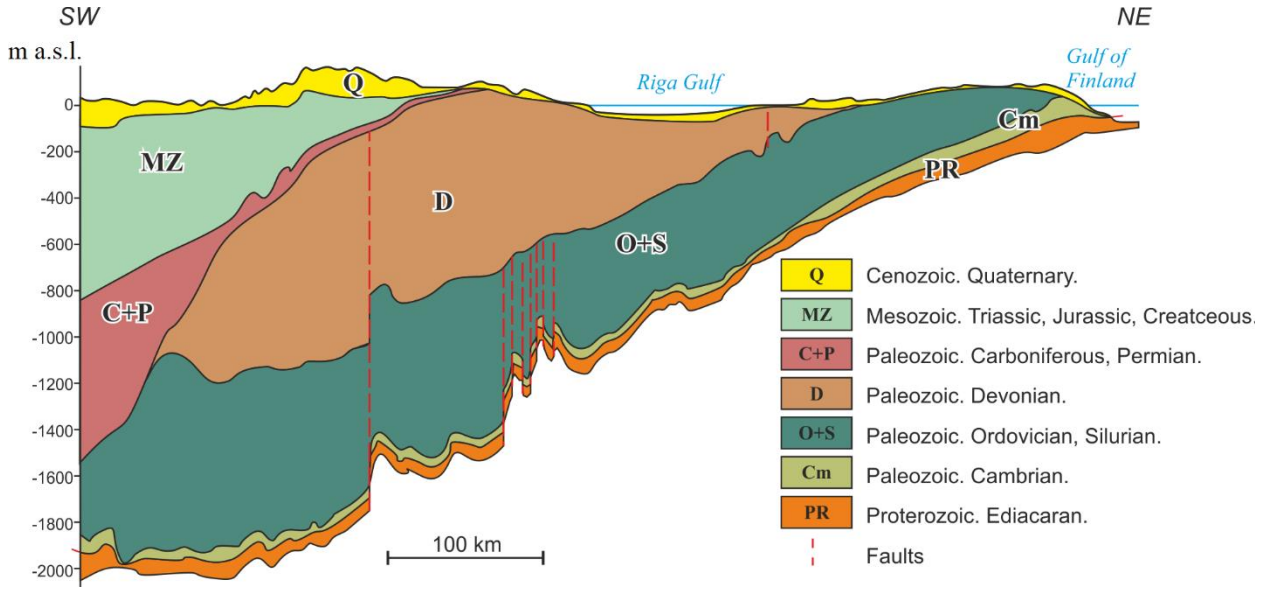

Figure 5. The geological cross-section of the Baltic countries (adapted from Krūmiņš et al. 2021).

The aquifers corresponding to the Devonian Era cover the Southern and Eastern parts of Estonia, all of Latvia, and the greater part of Lithuania (Sliaupa et al. 2008). The **Middle-Lower Devonian** aquifer system consists of Pärnu, Rēzekne, and Tilžė stages (first belonging to Middle Devonian but the latter two – Lower Devonian). They consist of fine-grained weakly cemented sandstones and siltstones interlaid with clay and dolomitized sandstone (Perens, Valner 2007). The thickness of the system is up to 200 m and ranges between 100 to 800 m in depth, with decreasing depth northwards (Babre et al. 2016; Levins et al. 1998). Throughout the whole territory of BAB the Middle-Lower Devonian is overlain by the Narva regional aquitard consisting of siltstone, marl, clay, and dolomite reaching the thickness of 90 m (Perens, Valner 2007).

The Amata and Gauja (Šventoji), Burtnieki, and Aruküla stages are considered The Middle Devonian in Estonia, but the Upper Devonian and Latvia and Lithuania, are located South of the Häädemeeste – Mustvee line (Perens, Valner 2007; Babre et al. 2016; Kitterød et al., submitted). The aquifer system consists of clay interlaid sandstones and siltstones. The upper part of the Amata stage forms an aquitard between **Middle and Upper Devonian**, and various local aquitards are believed to be present due to a high clay rock content in the strata (Perens, Valner 2007). The Upper Devonian series (Frasnian stage) consists of Dubņiki and Pļaviņas stages, containing fissured and karstified dolomites and dolomitized limestones (Perens, Valner 2007). The thickness of the aquifer reaches up to 300 m in southwest Latvia (Babre et al. 2016). The Snetaja Gora Member located in the lower part forms an aquitard due to the contents of siltstone with clay interlayers. The Middle-Upper Devonian system is an important source of drinking water in Lithuania (Mokrik et al. 2009) and Latvia, where the Upper Devonian Famennian

structure is used in the Southwest, Upper Devonian Pļaviņas-Amula system in the central and eastern part, but the northern part – Upper and Middle Devonian strata (Dēliņa 2018)

The aquifer systems corresponding to the **Paleozoic and Mesozoic Eras** are characteristic of Southwestern Latvia and southwards (Babre et al. 2016). Together with the **Cenozoic**, the maximum thickness reaches 600 m in Southwestern Lithuania. The Permian aquifer system consists of carbonates and evaporates of 100 m thickness, and it is overlaid by the Lower Triassic aquitard reaching up to 250 m in thickness (Sliaupa et al. 2008). The Jurassic sequence consists of clays, sandstones, and limestones and is up to 120 m thick. It is sparsely distributed and contained in Southwestern and Western Lithuania (Kitterød et al., submitted). Glauconite sand and chalk form the 140 m thick Cretaceous sequence which in Southwestern Lithuania is overlain by the Cenozoic sequence consisting of terrigenous sediments. The Upper Permian and Cretaceous aquifer systems are a significant source of drinking water in Lithuania (Sliaupa et al. 2008; Kitterød et al., submitted).

The quaternary aquifer consists predominantly of glaciogenic and marine sediments (Babre et al. 2016). It is used as a source of drinking water in Riga, Tallinn, and Narva cities as well as 60% of the territory of Lithuania (Dēliņa 2018; Gosk et al. 2007; Holman et al. 2000; Karro et al. 2009). In addition, the quaternary groundwater is extracted from shallow wells (Perens, Valner 2007; Dēliņa 2018). Within the territory of Latvia, the thickness of sediments varies, from up to 10 m to up to 50 m in Rīga and Daugavpils (Levins et al. 1998).

3.1.2. History of groundwater monitoring in the Baltic countries

The historical development of hydrological monitoring is similar between the Baltic countries - the aim of monitoring has changed over the years from locating new drinking water sources and determining their quality to adhering to the EU Water Framework Directive, Nitrates and Groundwater directives (Levins et al. 1998; Retiķe et al. 2022; Arustiene 2011). Regular, state-wide groundwater monitoring started in the 1960s (Bikše, Retiķe 2018; Juodkazis 1994; Arustiene 2011) but the improvement of automatic measuring came after the year 2000 (Arustiene 2011).

Estonia

According to the Estonian Geological Survey archive (https://fond.egt.ee) data, the first large-scale monitoring started with the research of Cambrian and Silurian aquifer systems for rational exploitation of groundwater resources. The results provide data on yearly averages and maximal and minimal groundwater levels per season across the years 1956-1957. Local exploration took place earlier. Bi-monthly measurements of groundwater levels were taken in Rakvere city in 1946 when describing the overview of geology and hydrogeology of the city.

Systematic groundwater monitoring in Estonia started in the 1960s, and while in 1961 there were 215 monitoring wells, it increased to 760 by the year 1990. National monitoring stopped during the collapse

of the Soviet Union and the initial years of independent Estonia and was resumed in 1995 as the beginning of the current monitoring system (Estonian Environmental Agency 2022).

Latvia

The national network for monitoring was developed between the years 1953 and 1959 when the first systematic measurements took place. The network was not extensive though, consisting of 15 wells in 4 stations (Retiķe et al. 2022), but the monitoring was improved in 1976 by installing wells in all active exchange hydrological systems (called the "floor system"). With the financial difficulties during the 1990s, the monitoring network decreased, and only 300+ wells were used (Levins et al. 1998). The improvement of automatic measurements took place between 2010 and 2012, and in the year 2019, bi-daily measurements supplemented by manual reference measurements took place in 301 wells of 74 stations. The digital groundwater monitoring database was created in the 1990s by the Latvian Geological Survey (subsequently Latvian Environment, Geology, and Meteorology Center) (Retiķe et al. 2022).

Lithuania

The first groundwater observations took place in the year 1946, while a state-wide groundwater network was implemented in 1963. Subsequently, alterations and improvements in monitoring took place in the years 1995, 2001, and 2005. The wells are divided into stations, and the wells within these stations can be individual or multiple. In some stations, the wells are nested, i.e., monitoring separate aquifers. Since 2005, daily groundwater levels are measured in 75 wells. A higher density of monitoring wells is located in the direction of Poland, but overall, all of the groundwater bodies of the country are monitored. (Arustiene 2011).

3.2. Data

The groundwater level data

The groundwater level data of all three Baltic countries was obtained from the responsible organizations – Latvian Environment, Geology and Meteorology Center, Estonian Environment Agency, and Lithuanian Geological Survey. The observation frequency of the level data table was varied – from a couple of entries per year to bi-daily, also within the same well. The start and end dates were varied (see Table 2), and multi-year missing data periods were present. The groundwater level data was accompanied by data of the respective stations (if applicable), coordinates, well elevation, respective horizon, well screen, and depth.

Table 2. The characteristics of the groundwater level dataset.

	Number of monitoring wells	Time interval	Number of entries	The beginning of daily/ bi-daily logging
Estonia	1149	1941-2018	1.49 million	2009-2014
Latvia	612 (74 stations)	1959-2019	1.68 million	2009-2014
Lithuania	266 (65)	1960- 2020	813 638	2004-2005

The raw ground water level time series contained various errors and was pre-cleaned using the 4-eye principle (one person initially corrects, another controls/reviews the results, makes suggestions) and expert judgment to increase the fitness for modeling (Retike et al. 2022). Alterations included correction or deletion of visible outliers and systematic errors. Well hydrographs that were unfit for modeling (for example, the dataset was only a year long) were eliminated. The encountered errors were either due to automatic logger problems (its failing or mismatch with manual reference measurements), visual, short-term water abstraction impacts, or human error registering levels, as well as other unclassified problems. The author of the thesis corrected 104 wells of Latvia and controlled 55 wells of Lithuania and 71 wells in Estonia.

For clustering that needs a uniform timestep, the groundwater level time series were imputed. MissForest (Stekhoven, Buehlmann 2012) gap-filling approach imputed time-series of 283 wells (see Figure 6) from the pre-cleaned data set having automatic logger daily or bi-daily data between January 2011 of 2011 and May 2019 of 2019 were obtained from *MSc* Jānis Bikše within the project “Spatial and temporal prediction of groundwater drought with mixed models for multilayer sedimentary basin under climate change” (LZP-2019/1-0165). The imputing was done with *missForest* (v. 1.4.) package for *R* programming language (Stekhoven, Buehlmann, 2012).

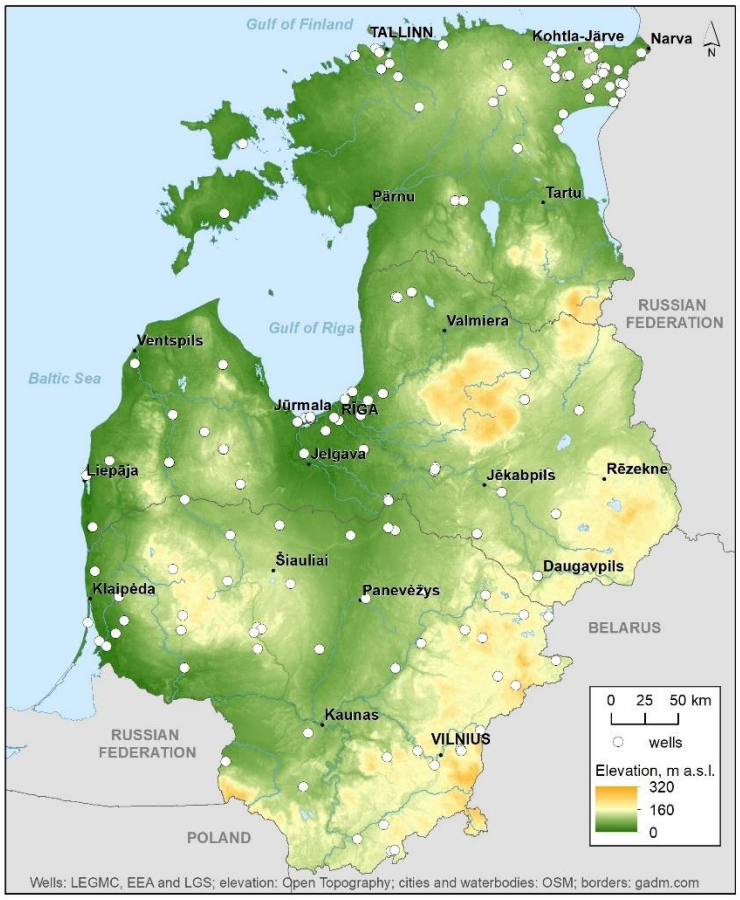

Figure 6. The location of observation wells.

Meteorological data

The daily precipitation and temperature data (mean, max, average) was obtained from European Climate Assessment & Dataset E-OBS gridded data ensemble (v. 24.0e). The dataset covers Europe with a grid size of 0.1 arc degrees. The temporal extent of the dataset is from January 1950 to June 2021. The dataset with daily time series was created as an interpolation of ground-level measurements from meteorologic stations (Cornes et al. 2018). The number of meteorological stations corresponding to the Baltic countries is visualized in Table 3.

Table 3. The station count of the three countries in the years 1951 and 2000 (TG – mean temperature, RR –precipitation) (European Climate Assessment &Dataset 2022).

	Estonia		Latvia		Lithuania	
	TG	RR	TG	RR	TG	RR
Stations in 1951	21	22	27	28	5	6
Stations in 2000	25	25	23	87	6	73

The potential evapotranspiration (PET) was calculated based on the Hargreaves-Samani equation (Hargreaves, Samani 1985). The equation was improved by calibrating the empirical coefficient C for local properties according to Knapp et al. (1980) and supported by Samani (2000). The calculations were done using daily maximum and minimum temperatures as well as daily precipitation sum from EOBS gridded dataset (Cornes et al. 2018).

Environmental descriptors

To find links between the model parameters, the groundwater level pattern, the model that best replicates groundwater level, and the environment, four groups of system characteristics - climatic, geological, river basin, land cover and topographic - were prepared for each observation well according to Haaf et al. (2020) (see Figure 7). The environmental descriptor data was obtained from *MSc* Jānis Bikše and *MSc* Konrāds Popovs within the project "Spatial and temporal prediction of groundwater drought with mixed models for multilayer sedimentary basin under climate change" (LZP-2019/1-0165).

The **climate parameters** include mean annual precipitation, temperature, and evaporation; aridity index; rainfall seasonality index; the mean precipitation of the driest, wettest, warmest, and coldest quarters; the ratio between coldest and warmest as well as driest and wettest quarters. They were calculated for the period of imputed groundwater levels using the EOBS gridded dataset (version 24.0e) (Cornes et al. 2018) for the respective cell centers. The Aridity index was calculated according to the FAO-UNEP bioclimatic index (e.g., Boschetto et al. 2010). The rainfall seasonality index was calculated according to Livada, Asimakopolous (2005). The mean precipitation of driest; wettest; warmest; coldest quarters was calculated according to O'Donnell, Ignizio (2012), and the quarters were defined as 12 consecutive sets of 3 months ($\sum_{i=1}^{i=3} P_i;\ \sum_{i=2}^{i=4} P_i; etc.$)

The **geological parameters** upper, lower level of the well screen; screen length; depth to groundwater; mean groundwater level; aquifer thickness; aquifer depth; the proportion of lithological classes in the cross-section was prepared using the geological cross-section and lithology data compiled from the Lithuanian Geological Survey, Estonian Land Board, and the European Social Fund project "Establishment of interdisciplinary scientist group and modeling system for groundwater research" (ESS 2009/8) of the University of Latvia. The cross-section data set contains information about the geological index, lithology (simplified to 12 classes), and the depth of the top and the bottom of the layers.

The **river basin parameters** the distance to the closest point on a stream and its height; the distance to the closest point of the river basin border and its height; elevation of the well; gradient to the stream were calculated using a Digital Elevation Model for the Baltic countries with resolution 10 m (compiled from DEMs from Estonian Land Board (Republic of Estonia Land Board 2022), Latvian Geospatial information agency (Latvian Geospatial information agency S.a.) and Lithuanian Geological Survey (provided upon request) and waterbody vector data for the three countries from Open Street Map (Open Street Map 2022), *R* programming language and GRASS GIS.

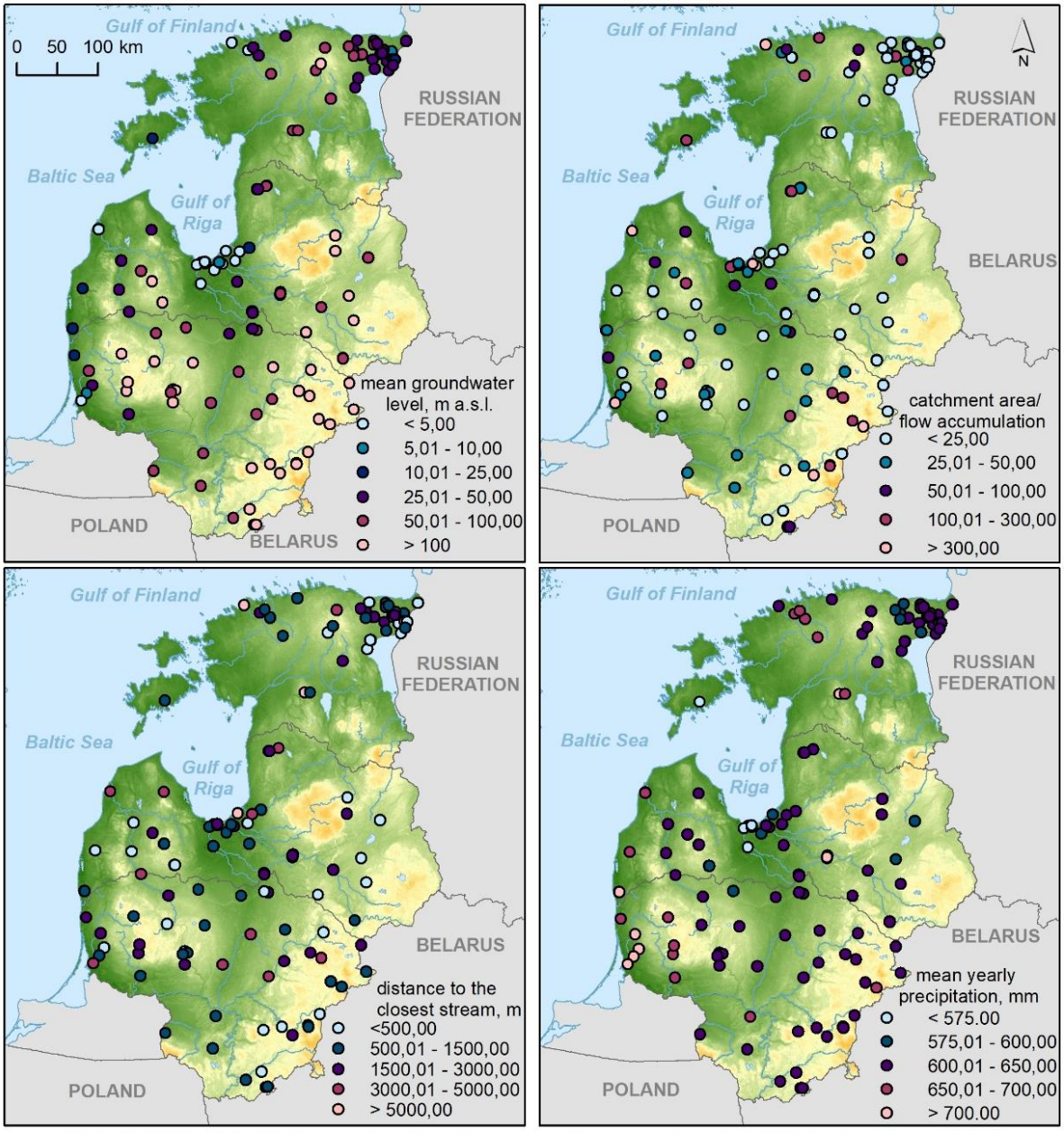

Figure 7. Example of environmental descriptors.

The **topography parameters** percentual area below observation well; the percentual position of observation well relative to the height range; mean topographic wetness index; mean slope; percentual concavity and convexity, and convergence; and the ratio between maximum catchment area and flow accumulation were calculated using the Baltic DEM (resolution 10 m) and GRASS GIS. The parameters were extracted over buffers around the imputed observation wells with a 500 m radius.

The **land cover parameters** the artificial surface area; the wetland area; the forest and seminatural area; the agricultural area and the waterbody area 500 m around the observation well was calculated from the Corine Land Cover data (European Environmental Agency 2022).

3.3. Methods

3.3.1. Time-series modeling using impulse response functions

When *N* stresses influence the groundwater head, the equation of the time series model is as follows (von Asmuth et al. 2008):

$$h(t) = \sum_{i=1}^{N} h(i)t + d + n(t) \qquad \text{(Eq. 1)}$$

where *h(t)* is the observed head at time *t*, *d* is the base level, *and n* is the residual series. The contribution of one stress to the head is characterized as (Collenteur et al. 2019):

$$h_{m\,(t)} = \int_{-\infty}^{t} S_m \tau \theta_m (t-\tau) d\tau \qquad \text{(Eq. 2)}$$

where S_m is time series of stress *m* and θ_m *is* impulse response function of stress *m.*

The impulse response function characterizes the groundwater level response to a unit of stress (such as precipitation). Generally, a scaled Gamma distribution function is used for modeling (von Asmuth et al. 2018):

$$\theta_p(t) = A \frac{a^n t^{n-1} exp\,(-at)}{\Gamma(n)} \qquad \text{(Eq. 3)}$$

where *A, a, n* are shape parameters of θ_p that are unknown and estimated, *t* is time.

However, the choice of the function depends on its capability to characterize the response accurately. The exponential function has only two parameters to estimate (Bakker et al. 2007):

$$\theta_p(t) = Aa\ exp(-at) \qquad \text{(Eq. 4)}$$

where A and *a* are shape parameters of θ_p that are unknown and estimated, *t* is time.

Alternatively, a four-parameter function is written as follows (Bakker et al. 2008):

$$\theta_p(t) = At^v e^{-at-b/t} \quad \text{(Eq. 5)}$$

where *A*, *v*, *a*, *b* – shape parameters of θ_p that are unknown and estimated, *t* - time.

The time series model consists of two parts: a deterministic part and a noise model (von Asmuth et al. 2002).

For the calculation of recharge, two approaches are commonly used as the transfer function – a linear model and a non-linear root zone model (Collenteur et al. 2021). The linear model assumes that the precipitation excess is a linear relationship between precipitation and evaporation:

$$R = P - fE_p \quad \text{(Eq. 6)}$$

where *R* – precipitation excess, *P* – precipitation, *f* - evaporation factor, E_p - potential evaporation.

A non-linear root zone model, on the other hand, assumes that the system can be modeled as two connecting reservoirs – interception and root zone reservoirs (see Figure 8), thus including short-term water retention of the soil in the calculations. The precipitation (P) firstly is retained in the interception zone until its capacity ($S_{i\ max}$) is reached while a certain amount of it evaporates (E_i). After exceeding the capacity of the interception zone, the remaining precipitation (P_e) reaches the second reservoir (the Root Zone), where part of it is evaporated by soil and plant processes ($E_{t,\ s}$), and the remaining water becomes groundwater recharge (R). Under certain conditions, a linear model combined with a more complex impulse response function can produce results similar to those of the root zone model. However, due to the linear model missing the capacity of soil moisture representation and hence introducing possible errors, the use of a nonlinear root zone model is advised (Collenteur et al. 2021).

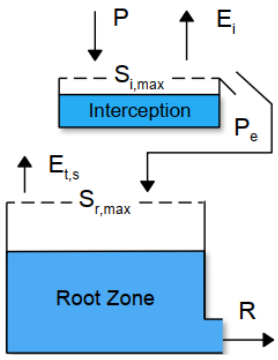

Figure 8. Non-linear recharge model (Collenteur et al. 2021).

Model calibration

The groundwater levels were modeled using time series models with impulse response functions, as implemented in the open-source package *Pastas* (Collenteur et al. 2019) in the *Python* programming language. They were modeled using daily time series of precipitation and mean daily temperature from the E-OBS gridded dataset (Cornes et al. 2018) and the pre-calculated potential evapotranspiration data from January 2011 to December 2017 as the drivers and historic groundwater level data for the calibration target. For each well, 4 models were created:

1. Linear model with a Gamma (3 parameters) function (LG);
2. Linear model with a four parameter function (L4);
3. Root zone model with an exponential function (RZE);
4. Snow model – root zone model with an exponential function using the daily mean temperature to determine the periods with snow cover (RZS). The RZS model follows the degree-day snow model based on Kavetski, Kuczera (2007) and determines the state of precipitation based on the daily mean temperature (Collenteur 2021b).

To determine if adding a trend increases the model fit significantly, the models were first solved without a trend and subsequently with a trend. The saved R^2 values corresponding to the calibration with and without trend were compared. The calibration with a trend was used if the value of R^2 for the calibration with the trend was at least 0.05 larger than that without a trend. The parameters were estimated with the non-linear least-squares method and were first estimated without a noise model and subsequently to foster a better parameter calibration.

Metrics R2 and RMSE were calculated using the *stats* model from the *Pastas* package, but the NSE and KGE were calculated using the *hydroeval* package in the *Python* programming language. They were calculated for both the calibration and validation periods (January 2011 to December 2017 and January 2018 to May 2019, respectively) and saved for subsequent model evaluation.

Model evaluation

The simulation cannot perfectly represent the true state of the system (Bennett et al 2013), and the simulation capability is evaluated for, e.g., comparison purposes. For evaluation of model performance, it is typically advised to calculate various metrics, since all of them separately characterize only a part of the dynamics, and thus give a non-objective interpretation of model effectiveness (Krause et al. 2005). A suggested and widely used practice is combining goodness-of-fit metrics with absolute error metrics and graphical techniques (Legates, McCabe 1999; Moriasi et al. 2007; Bennett et al. 2013).

The R^2 (coefficient of determination) is a goodness-of-fit measure that characterizes the proportion of variation explained by the model (Legates, McCabe 1999):

$$R^2 = \left(\frac{\sum_{i=1}^{n}(O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^{n}(O_i - \bar{O})^2}\sqrt{\sum_{i=1}^{n}(P_i - \bar{P})^2}}\right)^2, where \quad \text{(Eq. 7)}$$

O – observed value, *P* – predicted value, *n* – number of values. The values range between [0;1] and 1 indicates perfect fit, but 0 – no correlation.

A significant drawback of R^2 is its inherent capability to only characterize dispersion, therefore cases of constant over- or underprediction can still result in high R^2 values.

The RMSE is an absolute error metric that characterizes the error in units of measure (Legates, McCabe 1999) and ranges between [0; ∞] (Moriasi et al. 2015):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(O_i - P_i)^2}, where \quad \text{(Eq. 8)}$$

O – observation, *P* – prediction, *n* – number of values.

A drawback of RMSE is the tendency to overinflate the value in the presence of large outliers (Legates, McCabe 1999).

NSE (Nash-Sutcliffe efficiency), the normalization of mean squared error, is one of the most used metrics in hydrology (Gupta et al. 2009). NSE is a measurement of the variance of the residuals divided by the variance of observations. It characterizes the model as either better or poorer than using the mean value of the series (Nash, Sutcliffe, 1970; Schaefli, Gupta 2007).

$$NSE = 1 - \frac{\frac{1}{N}\sum_{t-1}^{N}(q_s^t - q_o^t)^2}{\frac{1}{N}\sum_{t=1}^{N}(q_o^t - \mu_o)^2}, where \quad \text{(Eq. 9)}$$

q_s^t – simulated value in time *t*, q_o^t – observed value in time *t*, μ_o – mean of observed values, *N* – number of values.

The output values can be in the range [-∞; 1], where 1 indicates perfect fit, and values below zero imply that the mean of the observed values gives better output than the simulation (Gupta, Kling 2011). The result of 0.5 indicates a satisfactory fit, but a good fit is considered from values 0.6 and above (Moriasi et al. 2007). NSE, the same as R^2, cannot objectively represent a systematic offset of values. In addition, NSE overemphasizes the impact of larger peaks, since the values are squared, and generally yield

smaller values in case of smaller amplitudes as compared to bigger ones (Krause et al. 2005; Schaefli, Gupta 2007).

KGE is a metric that is similar to NSE, but applies equal weights to the correlation, bias, and variability parameters, thus partly addressing the shortcomings of NSE (Gupta et al. 2009):

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2}\ , where \qquad \text{(Eq. 10)}$$

r – linear correlation between simulated and observed values, β – the ratio between mean simulated and mean observed value (the bias), α – the variability error of flow.

Apart from NSE, KGE does not have a specific benchmark, but the value -0.41 can be used when comparing the model against the mean of values. While similar in content, NSE and KGE are inherently different and can't be directly compared (Knoben et al. 2019).

The most appropriate model was chosen by both analyzing calculated metrics and using graphical techniques - visually observing the hydrographs ("good" if more than 60% of the variance was replicated in the simulation). The criteria of calibration and validation metrics for the model to be considered "good" are visualized in Table 4. The thresholds were determined either from literature studies or self-selected.

Table 4. The criteria for metrics for the model to be considered "good".

Metric	Value	Source
R2	>= 0.6	Brakkee et al. 2021
NSE	>= 0.6	Moriasi et al. 2007
KGE	>= 0.57	-
RMSE	<= 0.25	-

The model was considered "good" both if the metrics and visual evaluation were determined as "good".

3.3.2. Description of hydrographs

The level variability of groundwater hydrographs can be described in various ways. For example, Heudorfer et al. (2019) have devised a set of indices that describe hydrograph characteristics. They are aggregated in three general groups – the structure that describes the level variability considering the time domain, distribution that describes the variability irrespective of the time, and shape that describes the individual peaks and level declines. To explore and characterize the hydrograph diversity, hydrograph descriptors derived from Heudorfer et al. (2019) were used (Annex 1). The descriptors

were determined and calculated for the normalized hydrograph series. All calculations were done using *percentile* function in *NumPy* package implemented in *Python* programming language.

3.3.3. Clustering of hydrographs

To explore the hydrograph variability and group them according to their patterns, the clustering of the hydrographs was done using an agglomerative hierarchical clustering approach with Ward's linkage algorithm using the *scipy* package for *Python* programming language. To account for possible differences in time, the dynamic time warping approach was used when preparing the distance matrix (implemented in the *sktime* package for *Python* programming language). Before clustering, the time series were normalized – the values were recalculated into the range [0;1] using the *MinMaxScaler* function from the *sklearn* package for *Python* programming language. The results of clustering were inspected and the cutoff distance for hydrograph clustering was determined by the cluster count where a higher distance did not distinguish the patterns more clearly.

3.3.4. Correlations with environment descriptors

Connections between environment descriptors and goodness-of-fit metrics

The fit of the time-series model with IRF can be varied based on the hydrograph properties, which, in turn, are affected by the location of the observation well. To determine the factors that impact the model fit, the correlation and the significance between environmental and hydrograph descriptors and model metric NSE were calculated. The calculations were done for all hydrograph simulations and each model type separately. The correlation between environmental descriptors, hydrograph descriptors, and model parameters was determined using the *corr()* function from the package *pandas* for *Python* programming language. The significance levels were determined using the *pearsonr* function from the package *scipy* for the *Python* programming language. The correlation was determined separately for each model type and each parameter.

Linear regression was performed between the environment descriptor and the value of NSE. Linear regression was performed using *OLS* (ordinary least squares) function from the *statsmodels* package for *Python* programming language. The regression was performed for each parameter and descriptor separately. Both R^2 and *p* values were saved for later analysis.

Connections between environment descriptors and model parameters

Due to the specific fit to the well, the shape of the IRF characterizes the properties of the hydrological system, such as the response time to a unit of stress as well as the response to sudden and ongoing stress

(von Asmuth et al. 2018). To determine if there is a relationship and between the parameters and the environment descriptors how strong it is, correlation and linear regression analyses were performed for models with a “good” fit (both metrics and visually). The correlation analysis was performed for each model and each parameter separately. The linear regression was performed for each parameter of the model type and each descriptor separately.

4. Results

4.1. The description of hydrographs

To explore the selected groundwater hydrographs and group them based on their level variability, the hydrographs were normalized by recalculating the values in the range [0; 1] and subsequently clustered using the hierarchical clustering method with ward linkage and dynamic time warping. It resulted in 9 clusters at the dendrogram distance 9000 (see Figure 8).

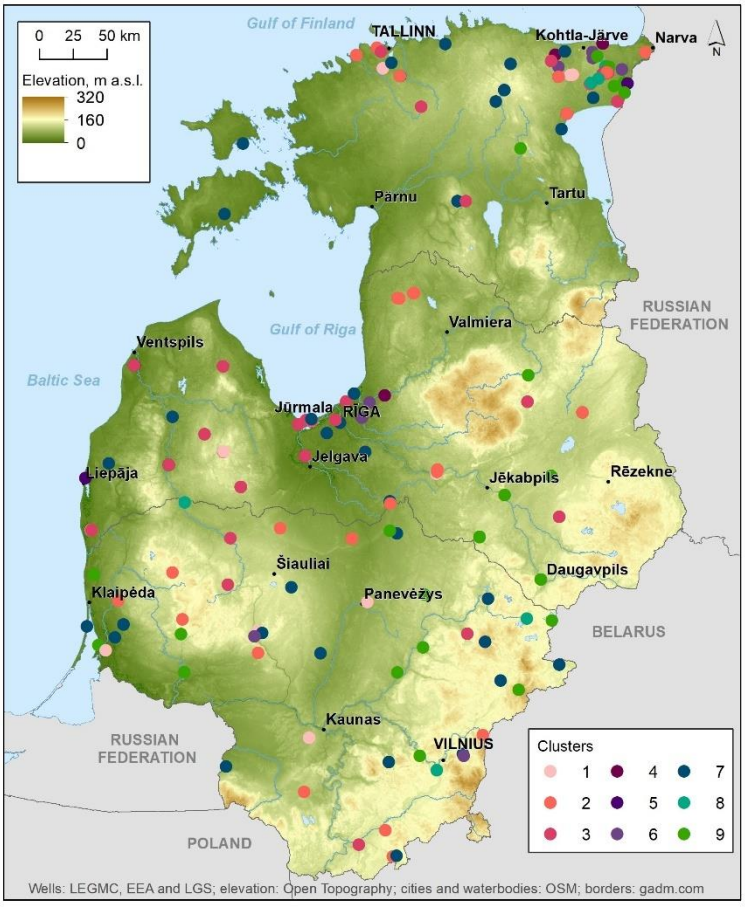

Figure 8. The spatial distribution of hydrograph clusters.

The clustering resulted in clusters with varying similarities. While the first cluster was made up of predominantly upper bound hydrographs (see Figure 9, upper left), the fifth cluster consisted of only three hydrographs with a somewhat stationary pattern and a steep upward trend subsequently (see Figure 9, lower left). Lower-bounded hydrographs with upwards peaks of various counts, width, and intensity (see Figure 9, upper right) were a characteristic pattern of the sixth cluster. All other clusters, however, had a mix of patterns, including sinusoidal hydrographs (see Figure 9, lower right) with a varying trend – either upward or downward, lasting the whole studied time frame of a couple of years.

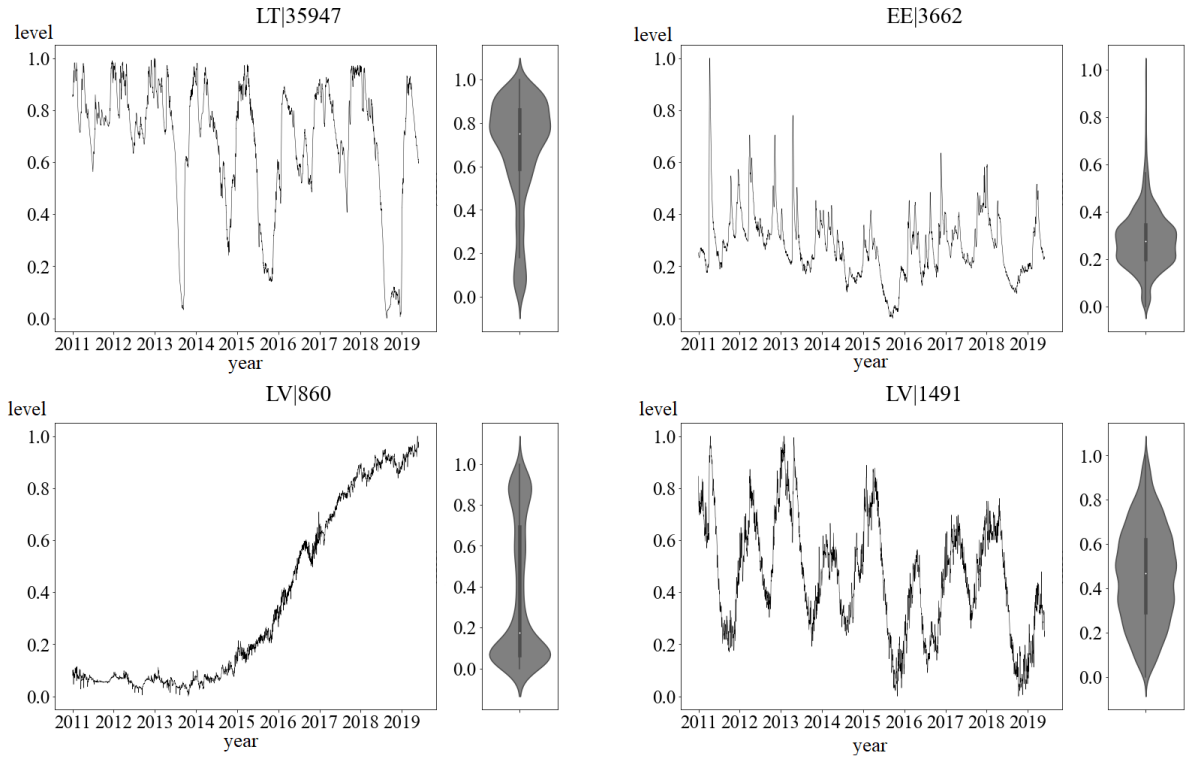

Figure 9. Examples of hydrographs.

Considering the assigned values for hydrograph characterization, the lowest medians correspond to lower bound hydrographs with long, short-lasting upward peaks of various recurrence, and the highest medians to upper-bounded hydrographs with downward troughs. However, in the cases of trended hydrographs, the median could be located anywhere within the range, depending on the steepness of the trend or its persistence. For stationery series with dominant seasonality, the median and mode were located around the 0.5 mark. The IQR was higher for either more stationary series or trended ones, while the lowest IQR values corresponded to hydrographs with pronounced upper or lower boundedness.

The selected time series had diverse patterns with varying structural and distribution characteristics. While there were instances where a set of hydrograph descriptors were characteristic of the same hydrograph pattern, it was not always the case.

4.2. Model performance evaluation

From the 283 time series, 20 were omitted due to a strong trend presence determined by visual evaluation. The remaining 263 groundwater level time series were modeled using time series models with impulse response functions and 4 different model structures – LG, L4, RZE, and RZS.

From the 1052 model simulations, 368 reached a “good” fit in the calibration period. The highest count of calibrations with a “good” fit was for the L4, but the lowest was RZE (see Table 5). Of the 368

models with a “good” fit in the calibration period, only 104 (28%) had a “good” fit in the validation period, corresponding to 55 well time-series, or 19% of the well time-series dataset. The highest number of models passing the validation requirements were L4 (32.4 %) followed by LG and RZS models with equal count (24.5 % each), leaving the RZE model with the lowest count of models with a “good” fit.

Table 5. Groundwater level time-series count for models that met the requirements of calibration, validation, and visual evaluation.

Model	Reached the metrics thresholds in the calibration period	Reached the metrics thresholds in the calibration period	Passed visual evaluation
LG	93	25	18
L4	118	33	25
RZE	73	19	12
RZS	81	25	17

Models passing the requirements for fit were subsequently subset according to their visual fit, resulting in 72 model simulations from 41 well time-series. The same as for calibration and validation steps, the highest model count of “good” fit was for the L4 model (34.7 %), but the lowest was RZE (16.7 %).

The median values of metrics R^2, NSE, and KGE in the calibration period were between 0.45 and 0.71. The minimum values were below zero, which in the case of NSE indicates that in such instances the mean of the time series could produce a better outcome than the simulated value. The same cannot be said about KGE, since none of the minimum values were below -0.41. In all cases, the median RMSE was below 20 cm. The resulting metrics in the calibration period of models with a “good” fit are visualized in Table 6.

Table 6. The metrics of calibration period for all model simulations.

Model	Metric	Minimum	First quartile	Median	Third quartile	IQR	Maximum
LG	R	-1.05	0.25	0.49	0.7	0.45	0.99
	NSE	-1.05	0.25	0.49	0.7	0.45	0.99
	KGE	-0.27	0.35	0.66	0.77	0.42	0.99
	RMSE	0.03	0.13	0.2	0.3	0.18	1.57
L4	R	-0.84	0.39	0.61	0.72	0.33	0.99
	NSE	-0.84	0.39	0.61	0.72	0.33	0.99
	KGE	-0.36	0.58	0.71	0.79	0.21	0.99
	RMSE	0.03	0.12	0.19	0.27	0.15	1.59
RZE	R	-0.89	0.21	0.45	0.64	0.43	0.99
	NSE	-0.89	0.21	0.45	0.64	0.43	0.99
	KGE	-0.35	0.31	0.58	0.71	0.41	0.99

	RMSE	0.04	0.13	0.21	0.32	0.19	4.27
RZS	R	-0.27	0.31	0.53	0.68	0.36	0.99
	NSE	-0.27	0.31	0.53	0.68	0.36	0.99
	KGE	-0.29	0.42	0.61	0.74	0.32	0.99
	RMSE	0.04	0.12	0.19	0.32	0.19	1.26

The L4 model had the highest median value for metrics R^2, NSE, and KGE (0.61-0.71), while also having the lowest value range. It is followed by the RZS model, and the LG and RZE model simulations had the lowest values of the metrics (0.45-0.66). The results indicate that the median time series corresponds to a good fit only for the L4 model simulations. In the case of RMSE, the lowest median value corresponded to the L4 and RZS model simulations (19 cm) (see Figure 10), however the absolute difference between the highest and the lowest median RMSE was only 2 cm.

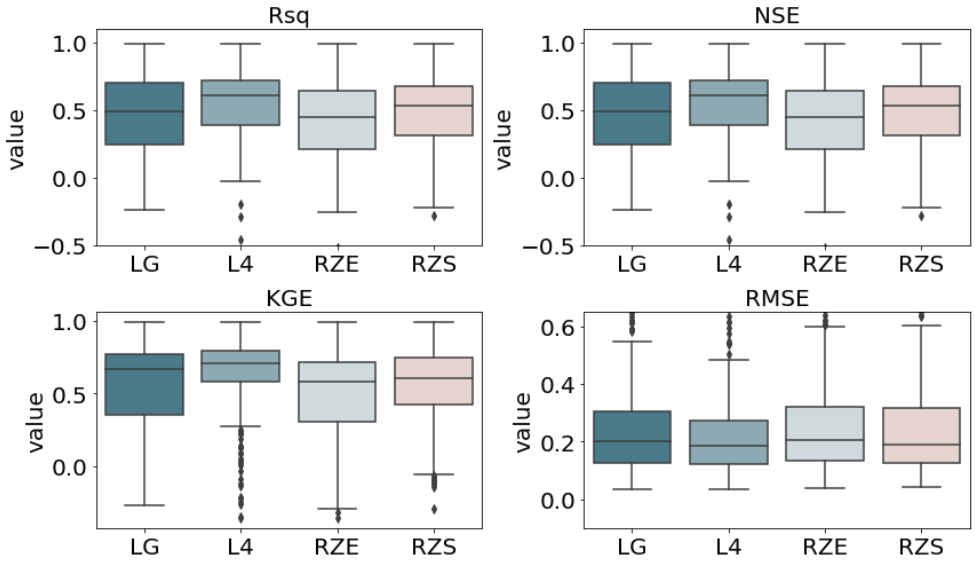

Figure 10. Boxplots of metrics for the calibration period of all simulations (zoomed in).

The median values for metrics R^2, NSE, and KGE in the validation period (see Table 7) were lower than the respective ones in the calibration period, with all of them exceeding 0, indicating that the simulated values of median time series of all model type datasets are a better prediction than the mean value of the time series. Together with the increased IQR as compared to the calibration results, they indicated that the fit in the validation period is poorer than in the calibration period. The respective median RMSE values have increased by up to 10 cm.

Table 7. The metrics of validation period for all simulations.

Model	Metric	Minimum	First quartile	Median	Third quartile	IQR	Maximum
LG	R	-1648.83	-0.42	0.06	0.47	0.89	0.92
	NSE	-28.05	-0.29	0.13	0.5	0.79	0.92
	KGE	-5.1	0.11	0.47	0.74	0.63	0.97
	RMSE	0.05	0.19	0.3	0.5	0.31	18.28
L4	R	-1365.61	-0.48	0.14	0.58	1.07	0.92
	NSE	-26.21	-0.21	0.3	0.63	0.85	0.92
	KGE	-5.1	0.38	0.64	0.79	0.41	0.96
	RMSE	0.05	0.18	0.27	0.48	0.3	6.75
RZE	R	-321.99	-0.39	0.06	0.42	0.81	0.92
	NSE	-323.36	-0.44	0.05	0.42	0.86	0.92
	KGE	-5.02	0.07	0.41	0.66	0.58	0.9
	RMSE	0.05	0.18	0.27	0.48	0.3	10.08
RZS	R	-62.95	-0.26	0.19	0.53	0.8	0.92
	NSE	-62.95	-0.34	0.19	0.53	0.87	0.92
	KGE	-4.91	0.20	0.47	0.66	0.45	0.93
	RMSE	0.05	0.18	0.26	0.44	0.27	4.45

The highest median value for R^2 corresponds to the RZS model (0.19) (see Figure 11), but the lowest to both LG and RZE models (0.06). However, the highest values of the NSE and KGE metrics correspond to the L4 model (0.4 and 0.64, respectively). The widest range between quartiles for R^2 was the L4 model, for NSE the RZE model, but for KGE the LG model. The lowest median RMSE corresponded to the RZS model with a difference of 4 cm with the highest value (0.3, LG).

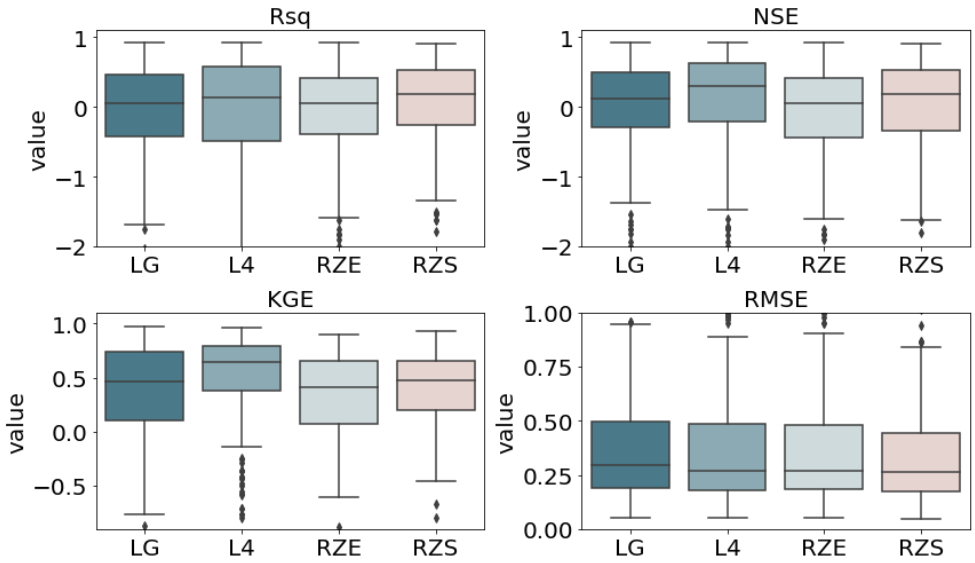

Figure 11. Boxplots of metrics for the validation period of all simulations (zoomed in).

263 well time series modeling with impulse response functions using 4 model types resulted in 72 different model instances with a "good" fit from 41 different well time series. L4 models count-wise were the most present in the final data set. The fit varied in both the calibration and validation period. Except for the L4 model simulations in the calibration period, the median time series fit of both the calibration and validation period was below the threshold of "good", but better than using the mean value of the time series (for NSE and KGE). The highest median fit in the calibration period was achieved with the L4 model, but both the L4 and RZS models had the highest replication capabilities (i.e., the highest values in the validation period).

4.3. The intercorrelations of environment descriptors

Multiple environmental descriptors can have a relationship - be correlated, which allows screening values for e.g., grouping (Singh et al. 2009; Sharma et al. 2015). In this case, correlation analysis was performed to decrease the number of descriptors. Intercorrelation of the descriptors, in this case, means that the descriptors will show a similar relationship to the studied variable. One of the colinear descriptors therefore can be eliminated to decrease the dimensionality of the descriptor data set. First, it was performed with the parameters within the same group (climatic, river basin, topographical, land cover) and subsequently on all descriptors together.

The river basin group consists of 7 descriptors (see Figure 12), namely distance to and elevation of the closest stream segment, distance to and elevation of the closest border of the catchment basin, elevation of the observation well, and ratios between well and stream segment elevation and well and catchment basin border elevation. High statistically significant correlations (> 0.9) (Singh et al. 2009; Sharma et al. 2015) ($p < 0.01$) were between parameters elevation of the closest basin border, well elevation, and elevation of the closest stream segment (see Figure 12) due to the similarity of well's location below the basin border and above the stream, and possibly due to all of them being located rather close by, considering the elevation in the scope of the whole study territory. The parameters elevation of the closest basin border and the elevation of the closest stream segment were excluded from the parameter list.

The topographic group consists of 8 descriptors (see Figure 12) – percentual position of the well within the height range, the percentual area below the well, slope, topographic wetness, percentual concavity, convexity, and convergence as well as the ratio between catchment area and flow direction. No descriptors reached a high correlation, however a moderate-high (> 0.6) (Singh et al. 2009; Sharma et al. 2015) correlation was between topographical wetness index and slope (-0.83, $p < 0.01$), concavity and convergence indices (0.72, $p < 0.01$) and convergence and convexity indices (0.69, $p < 0.01$). The topographic wetness index is calculated as the ratio between the upslope area and slope (Sorensen et al. 2006) which explains the inverse correlation. The descriptor slope was further excluded. The convergence index describes the relief's divergence and convergence (Jasiewicz 2022) while concavity and convexity describe the curve of the terrain. Since the latter two were calculated from convergence and have moderate-high correlations with high significance, they were further excluded.

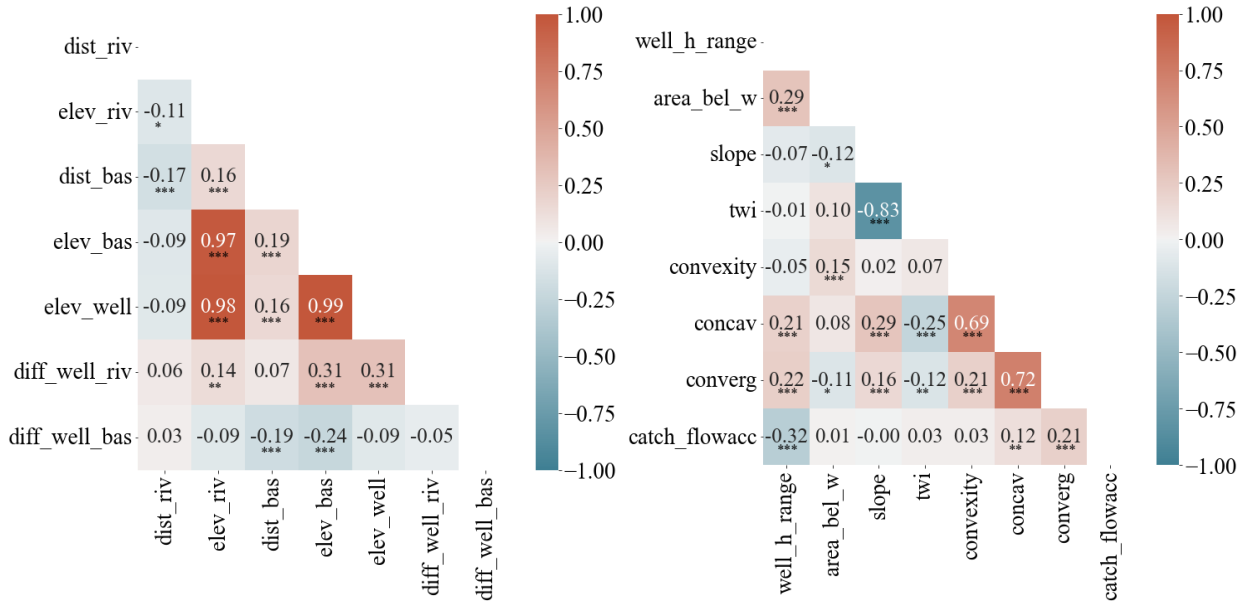

Figure 12. Environmental descriptors' intercorrelations of river basin group (left) and topographic group (right).

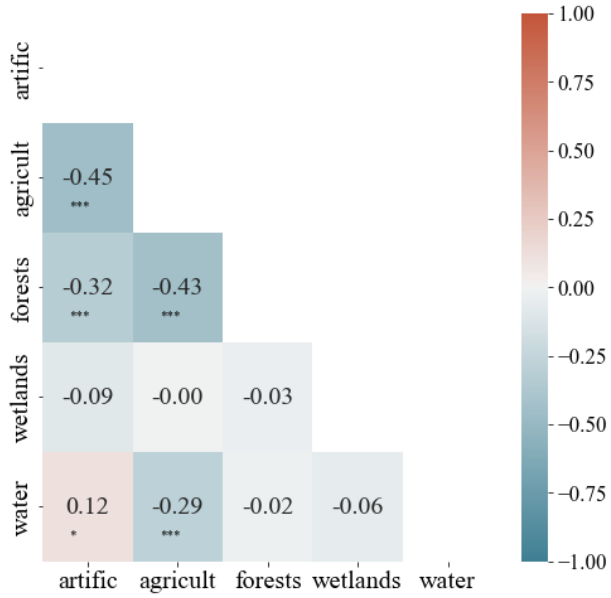

Figure 13. Environmental descriptors' intercorrelations of the land cover group.

No high correlations were found in the land cover group (see Figure 13) although the agricultural area class has a moderate inverse relationship with both artificial surfaces and forests and seminatural areas classes: -0.45 and -0.43 ($p < 0.01$), respectively. The relationship is most probably at the expense of each other since various wells are located in or near both cities and villages.

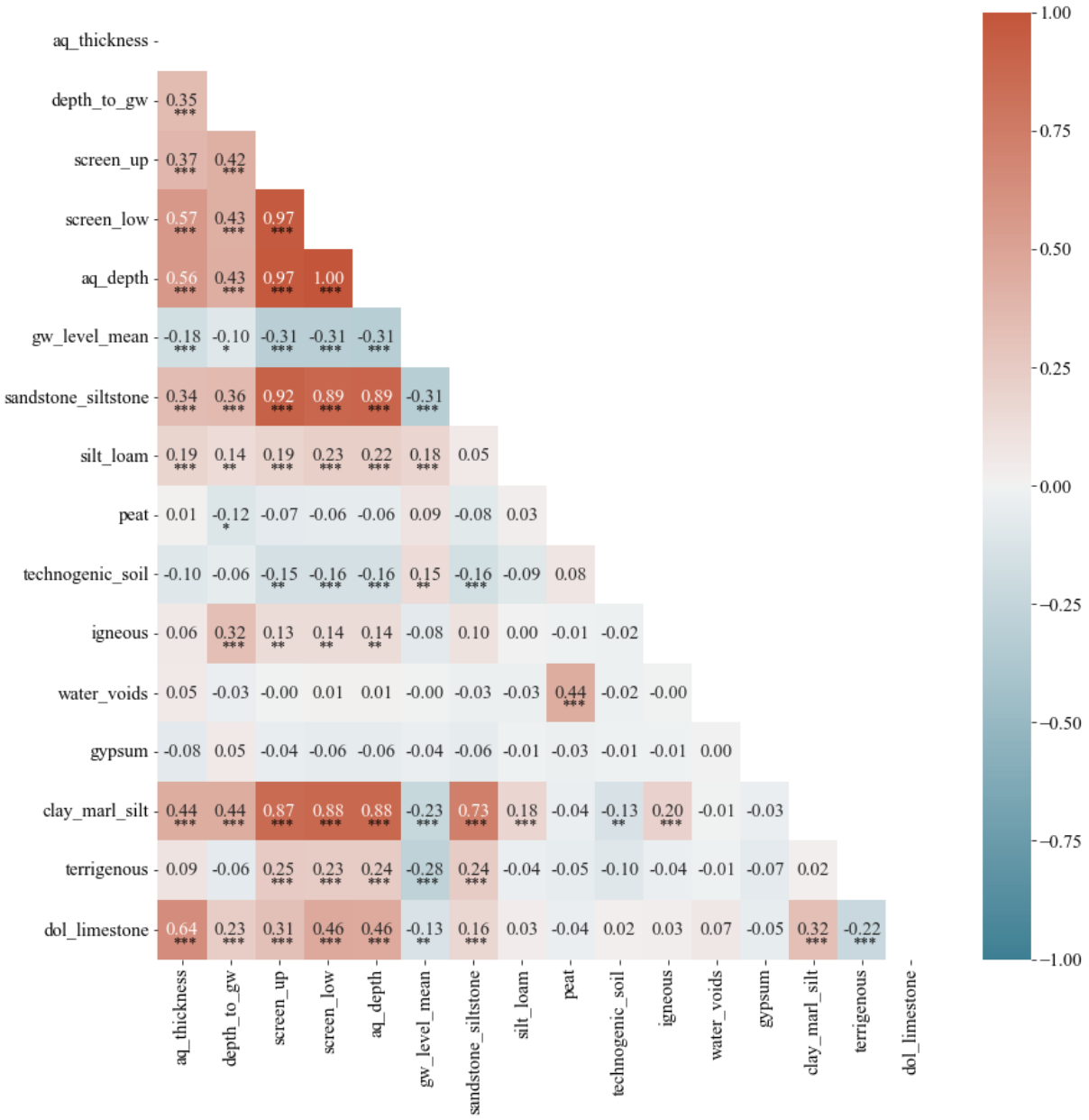

Figure 14. Environmental descriptors' intercorrelations of geology group.

The geology group consists of parameters describing the lithology and stratigraphy of the observation wells. Lithology parameters sand and gravel (terrigenous); clay, marl, and silt; dolomite and limestone were combined and summed in new groups due to their coexistence in strata and more straightforward interpretation. The parameters upper, lower screen depth and aquifer depth all intercorrelate with all correlations exceeding > 0.9 and high significance ($p < 0.01$) (see Figure 14), therefore only aquifer depth was retained. These parameters showed a moderately high correlation with sandstone and siltstone as well as clay, marl, and silt content in the profile, which could be attributed to both specific lithological conditions of the strata and their varying depths across the three countries, hence the correlation is neither high nor low and stays in the medium range. In addition, descriptors sandstone

and siltstone correlate with clay and silt content for the very same reason, but none of the lithological groups were excluded.

For climate descriptors (see Figure 15), mean precipitation in the coldest quarter highly correlates with mean precipitation in the driest quarter (1, $p < 0.01$) as well as the descriptors for wettest and warmest which implies that the respective seasons overlap. The rainfall seasonality index (SI) intercorrelates significantly (-0.92, $p < 0.01$) with both the ratio between coldest and warmest quarters and the ratio between the driest and wettest seasons and, since the intercorrelation of both is high (1, $p < 0.01$), they are excluded from the further study. In addition, parameters mean precipitation in the coldest and warmest quarters were excluded since rather high and significant correlations exist with rainfall seasonality index and average yearly precipitation (-0.71 and 0.8, respectively). The aridity index was excluded due to being calculated from both yearly precipitation and evapotranspiration.

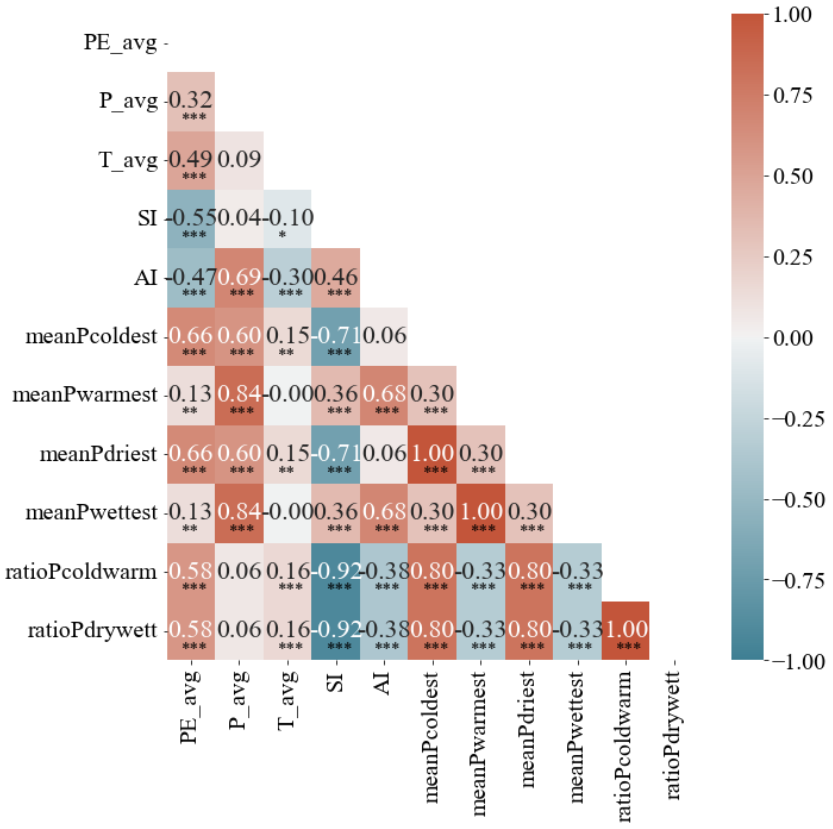

Figure 15. Environmental descriptors' intercorrelations of climate group.

Hydrograph descriptors differ from environmental descriptors in that they describe the groundwater level variations themselves. Four of the descriptors were assigned semi-quantitatively i.e., by visual observation and based on predefined ordered classes, but others quantitatively. Numerical descriptors median, first quartile, third quartile, first, and second mode all had significant correlations > 0.7 (see Figure 16), therefore only median was kept in the descriptor list for further study.

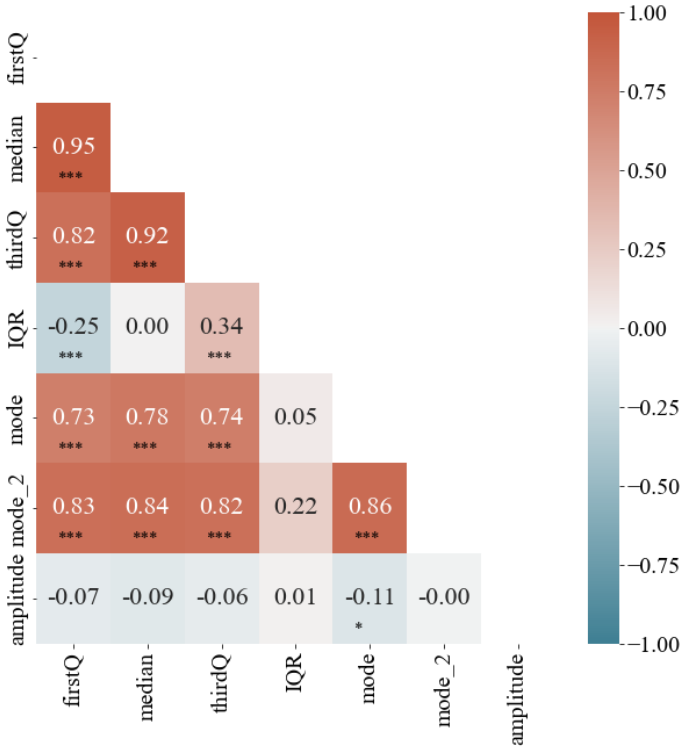

Figure 16. Hydrograph descriptors' intercorrelations.

An additional screening was performed to eliminate descriptors that have significantly high correlations with descriptors of other groups. A high correlation (0.98, $p < 0.01$) was between the well elevation and mean groundwater level descriptors since both were measured in absolute meters above the mean sea level. The descriptor well elevation was eliminated.

4.4. Relationships between the model fit and environment descriptors

The relationships between the obtained model fit and the environmental descriptors were determined by performing a correlation analysis and determining the significance of the correlations for each model type (LG, L4, RZE, RZS) separately. For this task, all simulations irrespective of the fit were used. The correlation results are shown in Figures 17 and 18.

A significant correlation with the descriptor IQR was characteristic of NSE of all model-type simulations (0.26-0.32, $p < 0.01$). In addition, three of the four models (L4, RZE, and RZS) had a statistically significant correlation with the descriptor median that describes the boundedness. The median descriptor had a rather low correlation and the direction of relationship varied (-0.12, $p < 0.01$ for RZE; 0.18 $p < 0.05$ for L4; -0.19 $p < 0.01$ for RZS). A higher IQR corresponded to both center bounded and trended time series for whom the trend was replicated quite well, therefore the resulting NSE value was high. While the correlation with the IQR indicates that a higher NSE value for all three model types was achieved for either sinusoidal stationery or trended time series, the L4 had higher NSE

values for upper-bound series, while the opposite can be said about the root zone models. The relatively low correlation coefficient with median could be due to various median values corresponding to diverse patterns that were replicated with varying success.

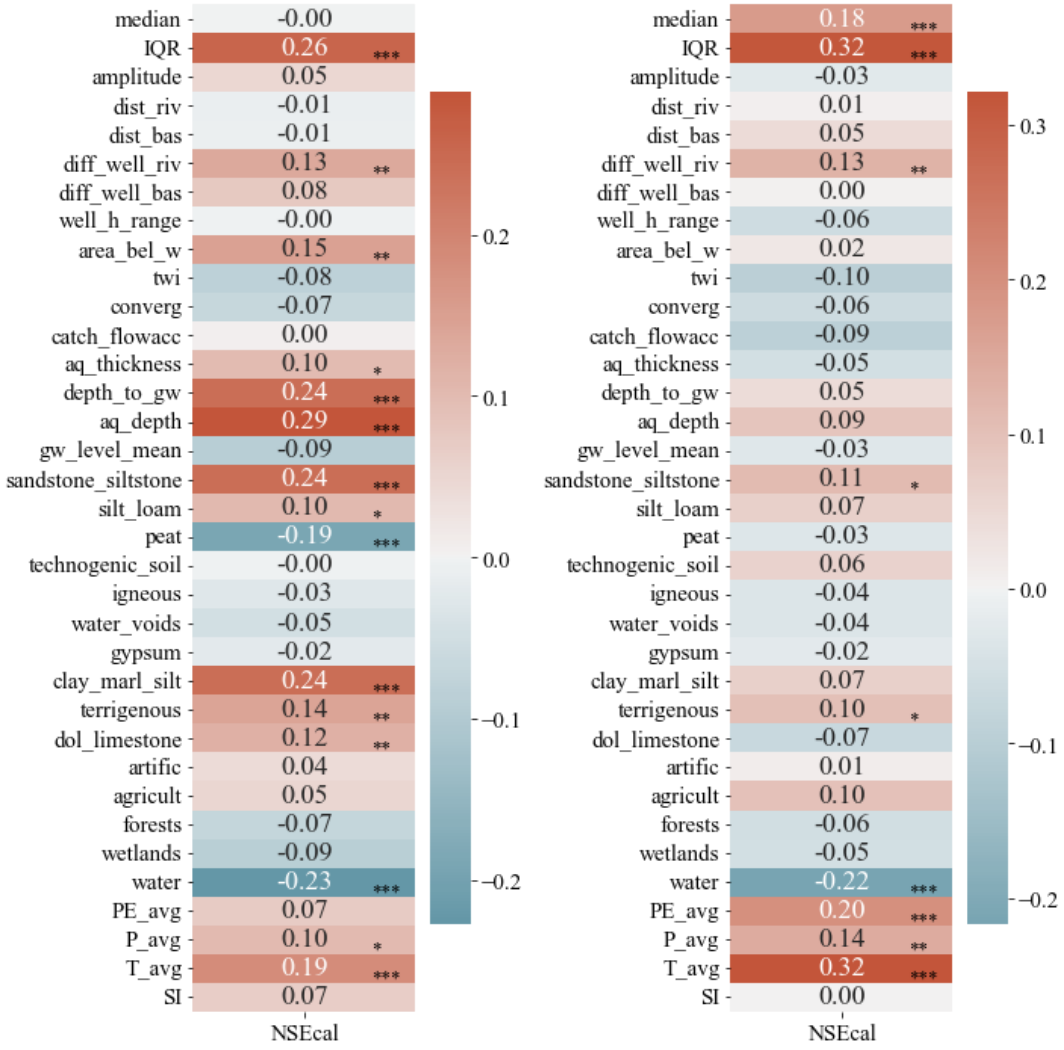

Figure 17. The correlation between best model's NSE value and environmental descriptors (left – LG; right – L4 model) * - $p < 0.1$; ** - $p < 0.05$; *** - $p < 0.01$.

The only statistically significant descriptor from the river basin group that was significant for NSE of all model types was gradient to the closest stream (0.13-0.16, $p < 0.05$). A higher gradient to the closest stream is consistent with wells higher on the local slope. Analyzing the topography parameters, the NSE value of LG simulations had a positive statistically significant correlation with the descriptor area below the well (0.15, $p < 0.05$), while both root zone models had a statistically significant correlation with the descriptor topographic wetness index (TWI) (-0.17, $p < 0.01$ and -0.14, $p < 0.05$, respectively). Considering that a higher area below the well indicates a location higher in the topography of the basin,

but a higher TWI value corresponds to locations with higher runoff and locations lower on the slope, a higher NSE value for all three model types (LG, RZE, and RZS) corresponds to observation wells located higher in the local topography.

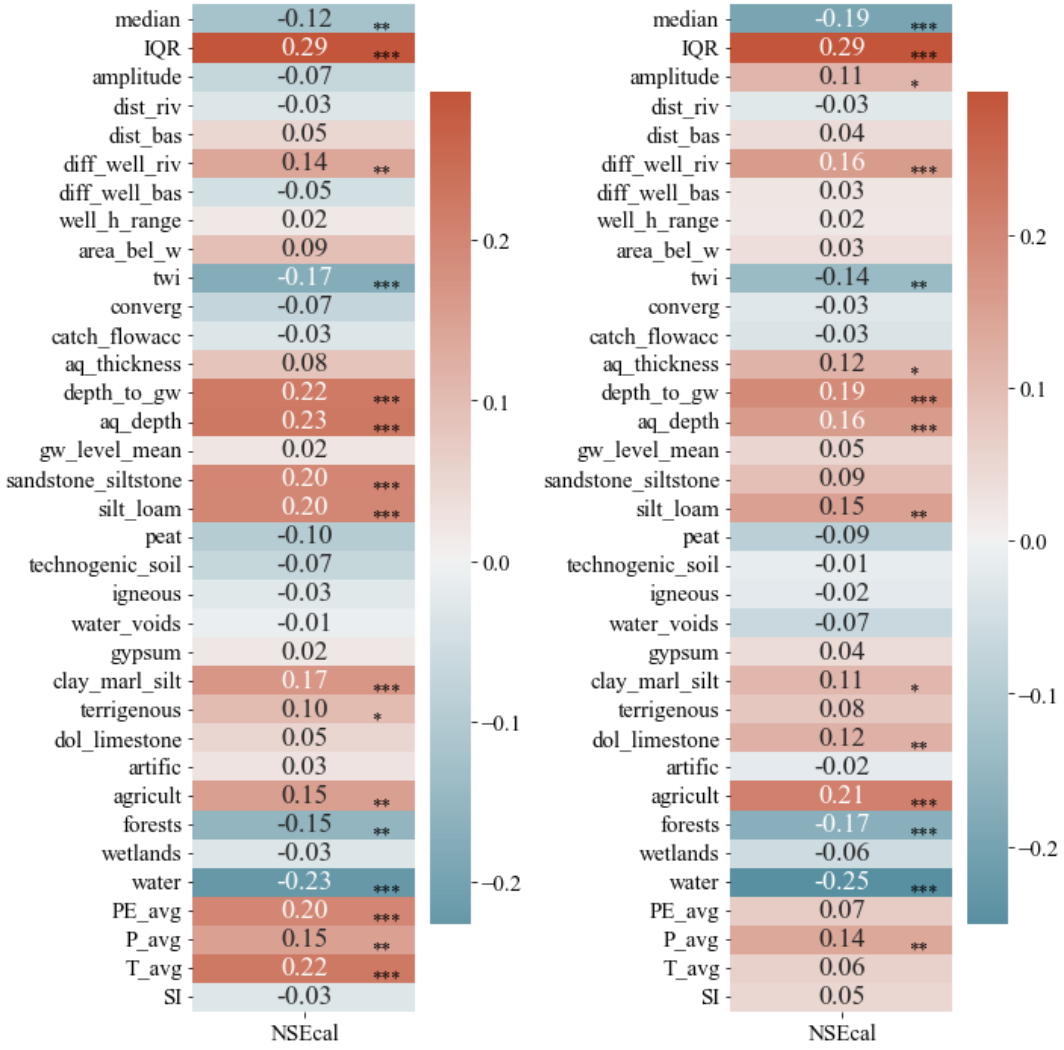

Figure 18. The correlation between best model's NSE value and environmental descriptors (left – RZE model; right – RZS model) * - $p < 0.1$; ** - $p < 0.05$; *** - $p < 0.01$.

From the geological parameters concerning stratigraphy, both the depth to groundwater level and aquifer depth had a statistically significant correlation with the NSE values of LG, RZE, and RZS model simulations (0.19-0.24, $p < 0.01$ and 0.16-0.29, $p < 0.01$, respectively). For LG and RZS, there was an additional statistically significant correlation with the descriptor aquifer thickness (0.10 and 0.12, $p < 0.1$, respectively), therefore for the model simulations of LG, RZE and RZS, the NSE value increases with a higher depth of both the groundwater and aquifer, while for LG and RZS a higher value of thickness corresponds to a higher NSE value.

In the case of lithology, three of the four models had statistically significant correlations with the descriptors sandstone and siltstone (LG, L4, RZE), silt and loam (LG, RZE, RZS), clay, marl and silt content (LG, RZE, RZS) and the terrigenous sediments content (LG, L4, RZE) in the profile. Of them, correlation of statistical significance for the descriptor sandstone and siltstone and terrigenous sediments was common for both linear models, but silt and loam as well as clay, marl, and silt content in the profile for both root zone models. The lithology impacts the permeability and hydraulic conductivity. With more permeable sediments, there is a higher infiltration and therefore recharge (Fitts 2002). In such a case, the recharge would mirror the precipitation pattern more, hence groundwater in terrigenous sediments would be more responsive than in clay-containing sediments. However, while it could be expected that the content of clay, marl, and silt as well as silt and loam content in the profile would have a negative relationship with the model fit, it was not the case in this study. It possibly indicates that the linear models perform better with higher presence of less permeable sediments.

Considering the land use descriptors, the NSE of all model simulations had a statistically significant negative correlation with the descriptor area of waterbody land class in 500 m radius (-0.22- -0.25, $p < 0.01$). In addition, both root zone models had a positive significant correlation with the land use class agricultural lands (0.15, $p < 0.05$ and 0.21, $p < 0.01$, respectively) and negative correlation with the land use class forests and seminatural areas (-0.15, $p < 0.05$ and -0.17, $p < 0.01$, respectively), with both descriptors intercorrelating.

While the descriptor average yearly precipitation had a positive significant correlation with NSE values of all model simulations (0.10 - 0.15), the parameter average yearly temperature had a statistically significant correlation coefficient with three (LG, L4, RZE, 0.19-0.32, $p < 0.01$). Average yearly evapotranspiration had a statistically significant correlation with NSE values of two model types (0.2, $p < 0.01$ for both L4 and RZE). All correlations were positive, meaning that a higher NSE value is expected at locations with higher precipitation, evapotranspiration, and temperature. Yearly average potential evapotranspiration depends on both the precipitation and the temperature, and these descriptors are affected both by the elevation and the distance from the sea (Jaagus et al. 2010), none of the models had a statistically significant correlation with the Seasonality index that spatially varies based on both the elevation and the distance to the sea.

To determine the strength of the relationship between the NSE value and the environment descriptors, a linear regression analysis was performed. While count wise there were a lot of statistically significant relationships from all descriptor groups, all the R^2 coefficients were low, with the highest value of 0.1 between the NSE of L4 model simulations and descriptor IQR (annex 1).

Four environment descriptors had a statistically significant R^2 coefficient with the NSE value of all model type simulations: IQR (0.066 – 0.1), average yearly precipitation (0.01-0.023), the gradient to the stream (0.016 – 0.025), and the area of water body land use class 500 m around the well (0.047 – 0.063). Descriptors for whom the R^2 value exceeded 0.05 for at least two models were from the groups of lithology, waterbody presence, stratigraphic and topographic make-up of the observation well location.

The NSE value of the model simulations for all four model types had statistically significant correlations with various environmental descriptors, most evidently the lithology of the profile, the

topographical make-up of the observation well location (elevation, slope, presence of water bodies, and the gradient to them) and the depth the groundwater level. A higher NSE value is consistent with locations higher on the local topography and with higher average yearly precipitation and average yearly temperature.

4.5. Relationships between IRF model parametrization and environment descriptors

Parameter *a*, *n*, and *A* values of the impulse response function for each model type were correlated against the environmental descriptors of the four different model simulations with a "good" fit individually to determine if they have a physical meaning (a relationship with environmental descriptors) and how strong these relationships are (see Figures 19 and 20).

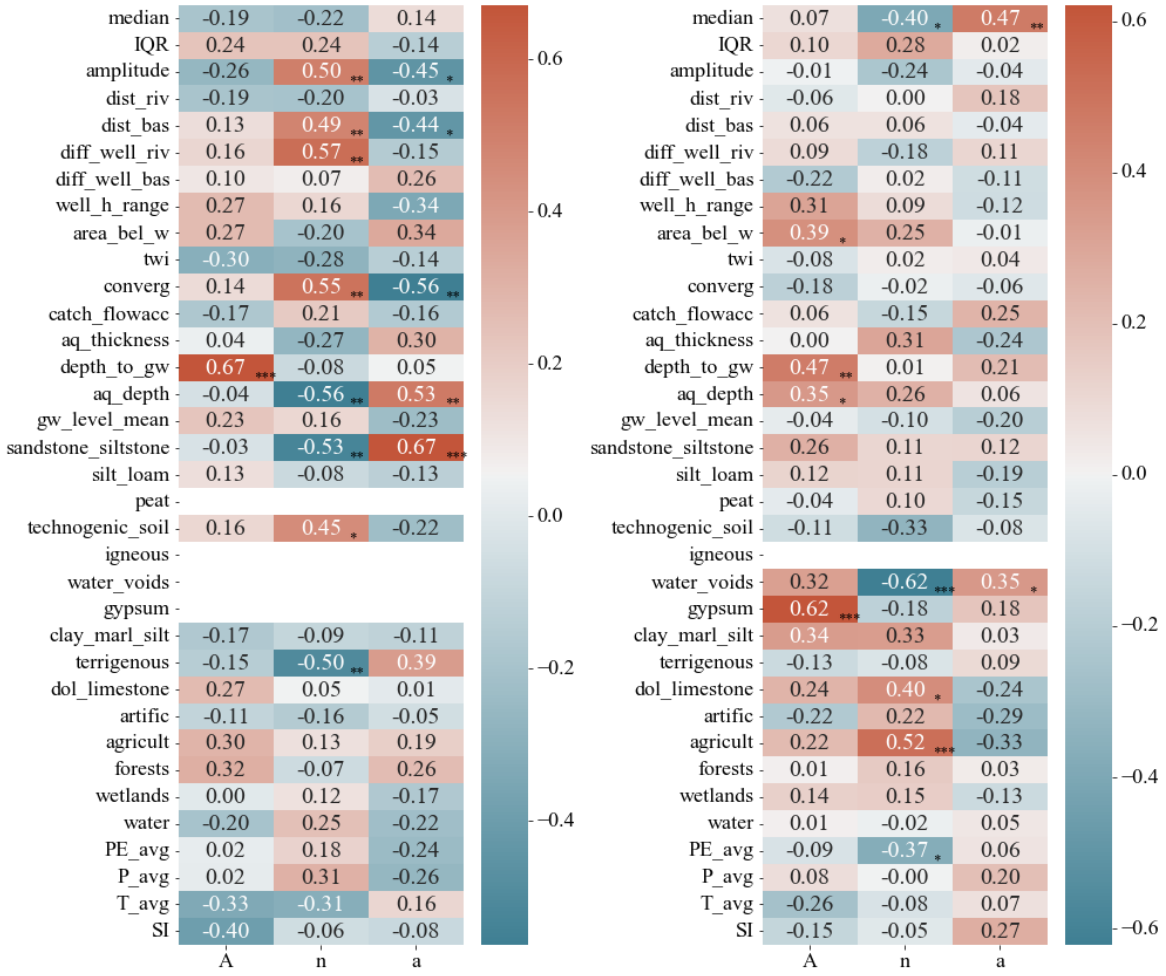

Figure 19. The correlations between the parameters of the model and the environmental descriptors (left – LG, right – L4) * - $p < 0.1$; ** - $p < 0.05$; *** - $p < 0.01$.

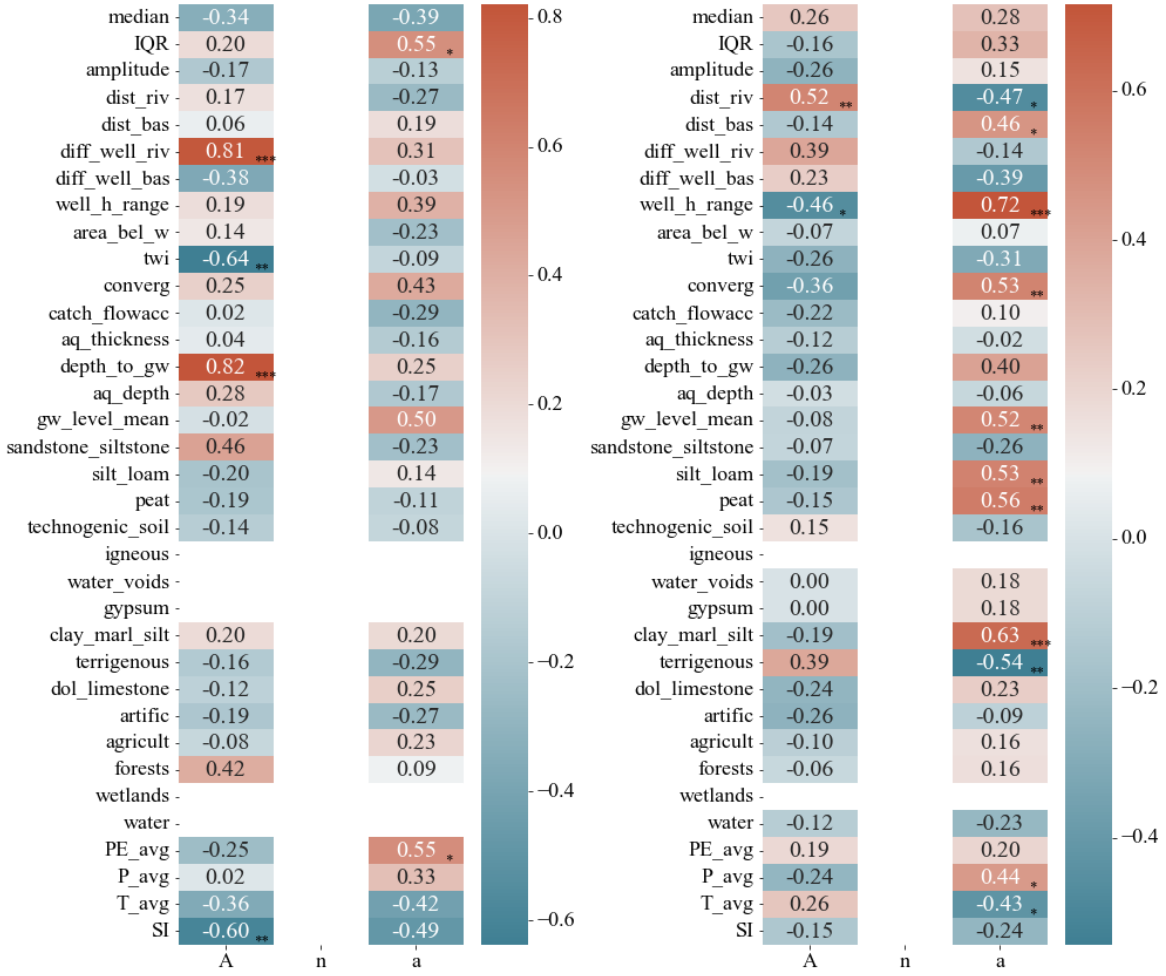

Figure 20. The correlations between the parameters of the model and the environmental descriptors (left – RZE, right – RZS) * - $p < 0.1$; ** - $p < 0.05$; *** - $p < 0.01$.

The parameter *A* of both linear models (LG and L4) had a statistically significant correlation with the environment descriptor depth to groundwater level (0.67, $p < 0.01$ and 0.47, $p < 0.05$, respectively). The parameter *A* of simulations from both root zone models (RZE and RZS, respectively) didn't have a common descriptor with a statistically significant value. However, parameter *A* of the RZE also had a statistically significant correlation (0.82, $p < 0.01$) with the descriptor depth to groundwater level. Additional correlations specific were present but were model-specific. For the L4, a correlation with the area below the well was present (0.39, $p < 0.1$), indicating that a higher value of parameter *A* is characteristic for wells higher in the catchment basin. In the case of RZE, two additional statistically significant moderate-high correlations with the descriptors gradient to the stream (0.81, $p < 0.01$) and topographic wetness index (-0.64, $p < 0.05$), both indicating that a higher value of *A* corresponds to wells located higher in the local topography. A similar conclusion can be drawn from another descriptor: the seasonality index in the Baltic countries follows the pattern of the relief (with an additional west to the east gradient in the case of Latvia and Lithuania).

In addition, a higher value of *A* for L4 is characteristic in locations with a deeper positioned aquifer as well as a higher presence of gypsum (0.35, $p < 0.1$ and 0.62, $p < 0.01$, respectively). The descriptors depth to groundwater level and aquifer depth somewhat correlate, however, the presence of gypsum is characteristic of only 1 well in the L4 dataset, located on both sides of the Latvian-Lithuanian border in the vicinity of Skaistkalne and Biržai.

For both root zone models, the parameter *n* equals 1. No environmental descriptor was significantly correlated with parameter *n* of both linear models' (LG and L4, respectively) simulations. However, various geologic descriptors of predominantly lithology had significant correlations with the parameter *n* with either of the linear models' simulations. For simulations of the L4 model, the value of *n* increases with increased content of dolomite and limestone in the profile (0.4, $p < 0.1$), but decreases with an increased presence of water voids (-0.62, $p < 0.05$) – strata affected by karst and suffosion. The simulations of LG and the parameter *n* value had a higher number of statistically significant correlations. Of them, a positive significant correlation between the parameter *n* value and the amount of technogenic soil in the profile (0.45, $p < 0.01$). Three descriptors had a negative relationship with the parameter *n* of LG, namely aquifer depth (-0.56, $p < 0.05$), sandstone and siltstone (-0.53, $p < 0.05$) and terrigenous sediment content in the profile (-0.5, $p < 0.05$). Therefore, the parameter *n* value is smaller in the locations with a higher presence of more permeable sediments and vice versa, as indicated by a negative correlation with more permeable lithological groups (water voids and terrigenous sediments) as compared to a less permeable one (technogenic soil). Both sandstone and siltstone have opposing permeability, but for this study, they were grouped.

Three additional statistically significant correlations were present for the parameter *n* value of the LG: with the environment descriptors distance to the closest basin border (0.49, $p < 0.05$), gradient to the closest stream (0.57, $p < 0.05$), and convergence (0.55, $p < 0.05$). Two of the descriptors indicate that the parameter *n* is higher in locations that are lower in the slope, i.e., further from the basin border and in a more converging area. The gradient to the closest stream, however, implies the opposite. The parameter *n* of the L4 had a statistically significant correlation with the descriptor average yearly evapotranspiration (-0.37, $p < 0.1$), therefore it could emphasize the impact of the regional relief, however, the values of the average yearly evapotranspiration across the three countries are more significantly impacted by the N-S and W-E gradient.

The parameter *n* of the LG had a statistically significant correlation with amplitude, implying that a higher value of *n* is characteristic of time-series corresponding to observation wells located higher on the slope (Haaf et al. 2020). In the case of the parameter *n* of the L4 simulations, a higher *n* value corresponds to lower median values, characteristic of hydrographs that are lower bound.

For the parameter *a* values of the LG simulations, the significant correlations with environmental descriptors were almost all the same as with the parameter *n* values (except for three), but inversed (i.e., negative correlation with parameter *a* if positive with parameter *n*) with the absolute value difference between 0.01 and 0.13. For the parameter *a* values of L4 simulations a similar pattern was present, although to a lesser degree – only two correlation coefficients were statistically significant for both parameters *a* and *n* values. In the case of both L4 and RZE, only two correlation coefficients were statistically significant and were unique to the model type.

Analyzing the statistically significant correlation between the parameter *a* values of the LG and RZS model simulations and the topographic descriptors, an inverse relationship is present. The correlation between the parameter *a* values of the LG model and descriptors distance to the basin border and convergence is negative (-0.44, $p < 0.1$ and -0.56, $p < 0.05$, respectively), but for the model RZS – positive (0.46, $p < 0.1$ and 0.53, $p < 0.05$). In addition, the parameter *a* of the L4 had a statistically significant correlation with the descriptors percentual position of observation well relative to the height range (0.72, $p < 0.01$) and distance to the closest stream (-0.47, $p < 0.1$). Therefore, a higher parameter *a* value of the RZS model is consistent with locations lower on the local slope (closer to the streams, further away from the basin border, having higher convergence), however, the parameter percentual position of observation well relative to the height range indicates the opposite.

While the parameter *a* value of the LG had a positive statistically significant correlation with the descriptor sandstone and siltstone content in the profile (0.67, $p < 0.01$), the L4 model *a* value had a positive statistically significant relationship with the descriptor water voids (0.35, $p < 0.1$). The RZS model parameter *a* values had a statistically significant correlation with various descriptors, namely the silt and loam (0.53, $p < 0.05$), peat (0.56, $p < 0.05$), clay, marl and silt (0.63, $p < 0.01$) content in the profile and negative correlation with the terrigenous sediments content in the profile (-0.54, $p < 0.05$). The parameter *a* value, therefore, in the case of RZS, is higher in the cases of less permeable sediments and lower with more permeable sediments. It also increases with a higher mean groundwater level, with correlates to the absolute elevation of the well, therefore a higher *a* value is characteristic of locations of higher absolute elevation. The parameter *a* value of the LG, however, had a positive statistically significant correlation with the descriptor aquifer depth (0.53, $p < 0.05$).

Considering the climatic descriptors, the parameter *a* of the RZE value had a positive statistically significant correlation with descriptor yearly average potential evapotranspiration (0.55, $p < 0.1$), while the L4 had a positive significant correlation with the parameter average yearly precipitation and negative with the average yearly temperature (0.44, $p < 0.1$ and -0.43, $p < 0.1$, respectively). Lower temperatures are consistent with both higher elevations and locations further to the north and east. The precipitation, however, has a more diverse spatial pattern, including based on the aspect of slope.

Three of the four models had one statistically significant correlation with one of the groundwater level variability descriptors. For the LG, it was significant with the descriptor amplitude (-0.45, $p < 0.1$). Considering the direction of the relationship and that a higher amplitude corresponds to wells located higher in the slope, the descriptor is congruent with the topographic descriptors distance to the closest basin border and convergence. In the case of the L4, a higher parameter *a* value corresponds to a higher median (upper bounded hydrographs), but for RZE – with a higher IQR value (more center bounded or tended hydrographs).

There were no common descriptor R^2 values that were statistically significant for all four model types. The value of parameter *A* of the LG model simulations predominantly had a statistically significant relationship with the descriptors sandstone and siltstone content in the strata (0.455, $p < 0.01$), convergence (0.312, $p < 0.05$) and aquifer depth (0.277, $p < 0.05$). The L4, however, had a statistically significant regression coefficient with the presence of gypsum in the strata (0.387, $p < 0.01$), depth to groundwater (0.216, $p < 0.05$) and the area below the well (0.149, $p < 0.1$). Similar to parameter *A* of

the LG model simulations, aquifer depth had a statistically significant regression coefficient with the parameter *A* value of the L4 model simulations, but the only statistically significant R^2 coefficient for the RZE model was the descriptor depth to groundwater level (0.451, $p < 0.01$). For the RZS model, both distance to the closest stream (0.272, $p < 0.05$) and the percentual position of observation well relative to the height range (0.216, $p < 0.1$) had statistically significant R^2 values.

While there were no common descriptors for both linear models, descriptors gradient to the well (0.329, $p < 0.05$), aquifer depth (0.316, $p < 0.05$), and convergence (0.307, $p < 0.05$) had the highest regression coefficients with the parameter *n* value of the LG model simulations, followed by lithology descriptors sandstone and siltstone and terrigenous sediment content and the distance to the closest basin border. For the parameter *n* values of the L4 model simulations, the most important descriptors were water voids (karst, suffusion) (0.329, $p < 0.01$) in the strata, agricultural land use class 500 m around the well (0.266, $p < 0.01$) and median (0.162, $p < 0.1$), followed by the dolomite and limestone content and average yearly evapotranspiration.

No descriptor was statistically significant for the parameter *a* of all model types. While the most important descriptors for LG were gradient to the closest stream (0.329, $p < 0.05$), convergence (0.307, $p < 0.05$) and aquifer depth (0.316, $p < 0.05$), the descriptors with the highest coefficient of the L4 model simulations' *a* value were median (0.222, $p < 0.05$) and water voids (0.123, $p < 0.1$). Sandstone and siltstone content in the profile was the most important descriptor for parameter *a* of RZE model simulations (0.445, $p < 0.01$), followed by convergence (0.312, $p < 0.05$) and aquifer depth (0.277, $p < 0.05$). The parameter *a* of the RZS model simulations, according to the linear regression results, had the highest regression coefficient values with the percentual position of observation well relative to the height range (0.513, $p < 0.01$), clay, marl and silt content in the profile (0.394, $p < 0.01$), and the peat content in the strata (0.312, $p < 0.05$) among other descriptors.

The parameters *n*, *A*, and *a* of the impulse response function had significant correlations and regression coefficients with various environment descriptors. Parameter *A* correlated to the descriptors depth to groundwater level and various descriptors of topography. The parameter *n* had significant correlations with various lithological classes, indicating that the parameter value differs based on the permeability of the strata. The parameter *a* had a statistically significant correlation with all descriptor groups apart from land use classes. While parameter *A* had higher values in locations consistent with a higher depth to groundwater, higher location of the local topography, and deeper aquifer, parameters *a* and *n* values were higher in the locations with a higher content of less permeable sediments.

5. Discussion

For this study, groundwater level time series between the dates of January 2011 and May 2019 were selected. This timeframe was selected due to the presence of daily to bi-daily data entries across all 283 wells. The groundwater level time series' gaps were imputed with the *MissForest* approach; therefore, the obtained metrics could differ in case the modeling was done on the gap-containing time series. In addition, the obtained metrics would be different if a longer period between the observations or simulation had been chosen - according to Berendrecht et al. (2003), the performance of the models increases when decreasing the interval of simulations given that peak response time is considered. On the other hand, a uniform time step ensured that there was no bias in the calibration or metrics' objectivity (Collenteur 2021a). The calibration period for modeling in this study was rather short – only six years. Zaadnoordijk et al. (2019) calibrated the models over eight years of data, while Collenteur et al. (2021) in their study had a ten-year calibration period. A longer calibration period could potentially increase the model fit and simulation capability.

The modeling outcome was possibly also affected by the pre-cleaning of the groundwater level time series. Outliers in hydrographs can be present both due to actual processes at the site and the human error in data handling (Peterson et al. 2017; Post, von Asmuth 2013). For this study, expert judgment was used to pre-clean time-series (Retiķe et al. 2022). While the task followed a set of repair actions based on the characteristics and the associated problem (such as the impact of pumping), low confidence in the respective actions was assigned to some of the problems, such as a change in the groundwater level pattern, for which the action of shifting or deleting was done. Hence some patterns could have been altered or deleted from the hydrographs that could contain information on relevant pattern changes and their possible causes, as described in e.g., Obergfell et al. (2019). Alternatively, time-series with a significant number of outliers or other disturbances in some studies are deleted (Zaadnoordijk et al. 2019) or more automated approaches to error detection are applied, such as far outlier removal or removal with time-series modeling (Brakkee et al. 2021).

The groundwater modeling was done using four different modeling approaches – a LG, a L4, a RZE, and RZS. A rather low count of the simulations fit the metrics thresholds - of the 1052 simulations, only 368 fit the metrics thresholds in the calibration period, but 104 of the 368 fit the validation period thresholds. The poor fit in the validation period could be attributed to changing spatial patterns in the validation period in the form of trends or groundwater level peak amplitude or width changes. A possible improvement of simulation count with a "good" fit could be increased by adding other impacting stresses, such as pumping or river level time-series as additional explanatory stresses (von Asmuth et al. 2007; Collenteur et al. 2019).

The resulting simulations were assessed visually to omit those with metrics corresponding to a "good" fit but poor replication of the groundwater level variability, resulting in 72 time series from 41 hydrographs. While such screening was valuable, it was subjective, therefore the replicability of such action is low (Barthel et al. 2022). Hydrographs of the time series that had sufficiently high metrics and their simulations replicated the observation variance well enough were mainly stationary or with a slight trend with a prominent seasonality. Time series that were lower-bounded hydrographs with

upward peaks of different occurrences, lengths, and persistence, and others with varying inter-annual patterns and system changes in the validation period did not reach a high model fit. In the study of Brakkee et al. (2021), the modeling approach gave poor results for wells with a deep groundwater level where inter-annual variability dominated, while Zaadnoordijk et al. (2019) emphasizes that a good model fit can't be reached where there is a strong influence from factors not accounted for in the explanatory series.

The highest count of the simulations that fit the metrics thresholds was from the L4, followed by the RZS model. Considering the median metric values, the best performing models were the L4 and RZS models. Berendrecht et al. (2006) suggest using the non-linear model for groundwater level modeling. Collenteur et al. (2021) concluded that the simulations obtained with the non-linear (root zone) model were better than those obtained with the linear model. The calibration of the models, therefore, could possibly be improved upon further.

The correlation analysis of the NSE of the four model type simulations and the environment descriptors showed that statistically significant correlations exist with various topographic and geological parameters, namely depth to groundwater level, terrigenous sediment content in the profile, gradient to the closest stream, and others. Similar statistically significant correlations were also found between the model simulation parameters and the environment descriptors. Some of the relationships between the environment descriptor and the NSE of model simulations could be due to the connection of pattern to a descriptor, for example, closeness to the nearest stream and long, short-lasting peaks (Giese et al. 2020) which in the case of this study were replicated poorly. However, these relationships have not been extensively studied, therefore the literature on this subject is scarce. While to the author's knowledge there currently are no studies describing the time series model with impulse response functions fit and parametrization relation to the environment of the observation well, a study on aiming to relate hydrograph characteristics to the environment was carried out by Haaf et al. (2020), the study based on which the environment descriptors were selected. In their study, multiple regression analysis results indicated that up to five descriptors were significant for models depending on confinement, and the most important descriptors were distance and gradient to the closest stream, depth to groundwater level, regional slope, and climatological descriptors, especially mean annual precipitation. In this study carried out by the author, geologic descriptors as well as topographic ones had the most important relationship (i.e., the highest coefficients) with the NSE and parameter values.

To decrease the number of environment descriptors and evaluate the relationship between studied value and environmental descriptors, Pearson correlation analysis was done within the parameter groups and subsequently between all descriptors. It resulted in a decrease of 15 descriptors, with the most from the climate and topographic descriptor groups, while the R^2 values of the linear regression between multiple environment descriptors and NSE, for example, were all low (< 0.15). A plausible explanation could be that the NSE value can be similar with different combinations of the environment descriptor values. Haaf et al. (2020) utilized not only Pearson correlation but also Spearman rank and distance correlation approaches and concluded that various descriptors have non-linear relationships. A similar study could potentially point out additional relationships, including non-linear ones. In addition, Haaf et al. (2020) performed a forward stepwise regression that iteratively builds a model based on various descriptors,

while for this study the correlation coefficient was determined between each parameter and each descriptor, while a more comprehensive model would consider multiple descriptors together.

Significant correlations were evident also with the hydrograph descriptors that were adapted from Heudorfer et al (2019). While various approaches of hydrograph classification can be employed (Barthel et al. 2022), for this study, the described indices were not calculated, but the concepts of the structure and distribution description were applied by ordinal class assignment for, e.g., primary, and secondary flashiness, or calculation of the IQR and median as a representation of the descriptors. While the calculated values seemed to describe the descriptors well, the ordinal class assignment was based on expert judgement that is hard to replicate (Barthel et al. 2022) and introduced subjectivity (Heudorfer et al. 2019). In addition, while an index value could describe a wide range of value distributions (Barthel et al. 2022), only three or four classes per descriptor were assigned.

6. Conclusion

The achievable fit of the time series models with impulse response functions has a large variability and depends on specificity of each observation well. Of the 1052 model simulations (for each of the four models and the 263 observation well time series), only 72 reached a “good” fit. The best-performing models determined by the number of “good” simulations and the median metrics were the L4 and RZS models. The hydrographs of the observation wells of a “good” fit correspond to mainly stationary series with or without a slight trend and dominating seasonality. The hydrographs that corresponded to a low fit had either prominent lower boundedness, pattern change during the validation period, or another variability that corresponds to other influences not used as input variables.

The fit of all four model types (LG, L4, RZE, RZS) as indicated by the goodness-of-fit metric the Nash Sutcliff efficiency (NSE) is higher for the observation wells with center-bounded or trended hydrographs, higher average yearly precipitation, and higher gradient to the closest stream, but lower with a higher presence of waterbodies in the vicinity of the observation well. Together with the model-specific correlations, a higher model fit corresponds to the locations higher in the local topography. The obtained coefficients of the linear regression analysis indicate that a significant heterogeneity of the environment descriptors can result in similar model fits.

The model parameters a, n, and A of the impulse response function have a physical meaning as determined by the statistically significant correlation and regression coefficients between the parameter values and the environment descriptors. While the parameter A values correlate to the depth to the groundwater level and various topographic descriptors, parameters’ a and n values predominantly correlate to various lithological classes’ presence in the profile.

7. Summary

The objectives of the work were to determine the obtainable fit of the groundwater level time series models with impulse response functions and analyze the relationships between various environmental descriptors and model fit and model parametrization, respectively. The groundwater levels of the Baltic countries were modeled with a designated package *Pastas* in *Python* programming language. Four modeling approaches were employed: a linear model with Gamma function, a linear model with four parameter function, a root zone model with an exponential function, and a root zone snow model. A "good" model fit was reached for 72 simulations that correspond to 41 groundwater time series out of the 283 modeled. The model fit varied based on both the groundwater level pattern as well as the uniformity of pattern throughout the studied time frame. The relationships between either model fit or the parameters of the impulse response function were determined by correlation and linear regression analysis. The model fit characterized by the NSE value of the model simulations had significant correlations with primarily the lithology parameters of the strata, the local topography around the observation well as well as the depth to the groundwater level. The statistically significant R^2 values between the environment descriptors and the NSE value of model simulations obtained from the linear regression analysis were low (< 0.15), which indicates that a significant heterogeneity can result in similar model fits. Notable correlations between parameters *A*, *n*, and *a* of the impulse response function and the environment descriptors were with descriptors depth to groundwater level, the amount of terrigenous sediments in the profile, aquifer depth, and others.

Põhjaveetaseme modelleerimine sademetepõhiste mudelite abil Balti riikides

Marta Jemeljanova

Kokkuvõte

Põhjavee puuringutes on vajalikud ühtlased aegread. Kuid seirejaamade põhjaveetaseme andmed sisaldavad tihti ebaregulaarseid ajalisi samme ning andmelünki, mis võivad ulatuda mitmetesse aastatesse. Põhjavee taseme sademetele reageerimise funktsioone (TFN-IRF) kasutavad mudelid on kasulikud tööriistad hankimaks päevaseid põhjaveetaseme andmete aegridu. Selliste mudelite parameetrite arv on väike ja arvutused kiired, kuid mudeli täpsus on siiski kõrge (von Asmuth et al. 2008). Sellist modelleerimise meetodit ei ole Balti riikide põhjaveetaseme mõõtmistele veel palju rakendatud. Samuti pole palju uuritud seirekaevude keskkonnaparameetrite mõju mudeli täpsusele.

Antud magistritöö eesmärgiks oli modelleerida Balti riikide seirekaevude põhjaveetaseme aegridade andmeid kasutades TFN-IRF meetodit ning hinnata antud meetodi sobivust põhjavee taseme modelleerimiseks. Selleks püstitati järgnevad uurimisküsimused:

1. Millist täpsust on võimalik saavutada kasutades TFN-IRF meetodit põhjaveetaseme ja meteoroloogiliste andmete modelleerimiseks Balti riikides?

2. Millised on seosed seirekaevude tunnuste ja mudeli täpsuse vahel?

3. Millised on seosed seirekaevude tunnuste ja mudelite parameetrite vahel?

Püstitatud uurimisküsimustele vastamiseks modelleeriti põhjaveetaseme aegridade andmeid Python-i programmeerimiskeele Pastas paketiga. Kasutati nelja modelleerimismeetodit: Gamma funktsiooniga lineaarmudel, neljaparameetrilise reageerimisfunktsiooniga lineaarmudel, eksponentsiaalfunktsiooniga juurtetsooni mudel ja juurtetsooni lumemudel. Aegridade ja uurimisalade tunnuste vahelised seosed leiti korrelatsioonide ja lineaarse regressiooni analüüsiga. Modelleerimiseks kasutati põhjaveetaseme andmeid kõigist Balti riikidest – Eestist, Lätist ja Leedust – aastatel 2011–2019, ning päevaste sademete ja õhutemperatuuride andmeid EOBS andmebaasist. Aurustumise aegread arvutati EOBS andmete põhjal Hargreaves-Samani võrrandi järgi. Tunnuste vaheliste seoste leidmiseks kasutati viie tunnusrühma (geoloogilise, topograafilise, vesikonna, kliima ja maakatte) andmestikku.

Saadud tulemused näitasid, et hea mudeli täpsuson saavutatav vaid 72 juhul 1052-st simulatsioonist, mis vastavad 41-le erinevale aegreale 263-st. Suurim arv „parimaid“ mudeleid saavutati neljaparameetrilise reageerimisfunktsiooniga lineaarmudeliga, kuid kõrgeim mediaansobivus saavutati mõlema juuretsooni mudeliga. Mudeli headus varieerus sõltuvalt konkreetse põhjaveekaevu taseme

kõikumiste mustritest. Valideerimisperioodil oli mudeli madala täpsuse oluliseks põhjuseks põhjaveetaseme kõikumiste mustrite muutumine.

Mudeli täpsust hinnati Nash-Sutcliffe indeksiga (NSE) - ning mudelisimulatsioonide täpsusel oli valdavalt oluline seos kihtide litoloogia, kohaliku topograafia ja põhjaveetasemega. Parem mudeli täpsus (kõrgem NSE väärtus) on iseloomulik topograafiliselt kõrgemal asuvatele kohtadele, kus aasta keskmine sademete hulk on suurem. Kuigi mitme keskkonnaparameetri ja NSE vahelised seoses olid statistiliselt olulised, olid kõik determinatsioonikordajad madalad (< 0,15), mis näitab, et seirekaevu tunnuste heterogeensus võib and sarnase mudeli täpsuse.

Acknowledgements

I would like to first thank all the supervisors: *Ph.D.* Alexander Kmoch, *MSc* Jānis Bikše and *MSc* Raoul Collenteur for their invaluable input throughout the whole development of the thesis, from the first ideas to the final alterations. In addition, I would like to thank *MSc* Konrāds Popovs for the preparation of the descriptor data and to *Dr.geol.* Andis Kalvāns for the ongoing support and introducing me to the topic of impulse response functions.

This study was supported by the Latvian Council of Science, project "Spatial and temporal prediction of groundwater drought with mixed models for multilayer sedimentary basin under climate change", project No. lzp-2019/1-0165.

References

Ahmad, S., Liu, H., Günther, A., Couwenberg, J., Lennartz, B. 2020. Long-term rewetting of degraded peatlands restores hydrological buffer function. Science of the Total Environment 749, 141571. doi: 10.1016/j.scitotenv.2020.141571

Ahmadzadeh, H., Mansouri, B., Fathian, F., Vaheddoost, B., 2022. Assessment of water demand reliability using SWAT and RIBASIM models with respect to climate change and operational water projects. Agricultural Water Management 261, 107377. doi: 10.1016/j.agwat.2021.107377

Alley, W.M., La Baugh, J.W., Reilly, T.E. 2006. Groundwater as an element in the hydrological cycle. In Encyclopedia of Hydrological Sciences (eds Anderson, M.G., McDonnell, J.J.). doi: 10.1002/0470848944.hsa153

Allen, D.M., Whitfield, P.H., Werner, A. 2010. Groundwater level responses in temperate mountainous terrain: regime classification, and linkages to climate and streamflow. Hydrological Processes 24(23), 3392–3412. doi:10.1002/hyp.7757

Amanambu, A. C., Obarein, O. A., Mossa, J., Li, L., Ayeni, S. S., Balogun, O., Oyebamiji, A., Ochege, F. U. 2020. Groundwater system and climate change: Present status and future considerations. Journal of Hydrology, 589, 125163. doi: 10.1016/j.jhydrol.2020.125163

Arnold, J.G., Srinivasan, R., Muttiah, R.S. and Williams, J.R. 1998. Large area hydrologic modeling and assessment Part I: Model Development. JAWRA Journal of the American Water Resources Association 34, 73-89. doi: 10.1111/j.1752-1688.1998.tb05961.x

Arustiene, J. 2011. Groundwater monitoring in Lithuania. In Groundwater Management in the East of the European Union (eds Nalecz T.). NATO Science for Peace and Security Series C: Environmental Security. Springer, Dordrecht, doi: 10.1007/978-90-481-9534-3_6

Apurv, T., Sivapalan, M., Cai, X. 2017. Understanding the role of climate characteristics in drought propagation. Water Resources Research, 53, 9304– 9329. doi: 10.1002/2017WR021445

Babre, A., Kalvāns, A., Avotniece, Z., Retiķe, I., Bikše, J., Popovs, K., Jemeljanova, M., Zelenkevičs, A., Dēliņa, A. 2022. The use of predefined drought indices for the assessment of groundwater drought episodes in the Baltic States over the period 1989–2018. Journal of Hydrology: Regional Studies 40, 101049. doi: 10.1016/j.ejrh.2022.101049

Babre, A., Kalvāns, A., Popovs, K., Retiķe, I., Dēliņa, A., Vaikmäe, R., Martma, T. 2016. Pleistocene age paleo-groundwater inferred from water-stable isotope values in the central part of the Baltic Artesian Basin. Isotopes in Environmental and Health Studies 52(6), 706-725. doi: 10.1080/10256016.2016.1168411

Bailey, R.T., Bieger, K., Flores, L., Tomer, M. 2002. Evaluating the contribution of subsurface drainage to watershed water yield using SWAT+ with groundwater modeling. Science of The Total Environment 802, 149962, doi: 10.1016/j.scitotenv.2021.149962

Bakker, M., Maas, K., Schaars, F., Von Asmuth, J. R. 2007. Analytic modeling of groundwater dynamics with an approximate impulse response function for areal recharge. Advances in Water Resources 30(3), 493-504. doi: 10.1016/j.advwatres.2006.04.008

Bakker, M., Maas, K., Von Asmuth, J. R. 2008. Calibration of transient groundwater models using time series analysis and moment matching. Water Resources Research 44, W04 420, doi: 10.1029/2007WR006239

Bakker, M. Schaars, F. 2019. Solving groundwater flow problems with time series analysis: You may not even need another model. Ground Water 57, 826-833. doi: 10.1111/gwat.12927

Barron, O. V., Crosbie, R. S., Dawes, W. R., Charles, S. P., Pickett, T., Donn, M. J. 2012. Climatic controls on diffuse groundwater recharge across Australia. Hydrology and Earth System Sciences 16(12), 4557-4570. doi: 10.5194/hess-16-4557-2012

Barthel, R., Haaf, E., Nygren, M., Giese, M. 2022. Systematic visual analysis of groundwater hydrographs: potential benefits and challenges. Hydrogeology Journal 30, 359–378. https://doi.org/10.1007/s10040-021-02433-w

Beck, H., Zimmermann, N., McVicar, T., Vergopolan, N., Berg, A., Wood, E.F. 2018. Present and future Köppen-Geiger climate classification maps at 1-km resolution. Scientific Data 5, 180214 doi: 10.1038/sdata.2018.214

Becker, A., Serban, P. 1990. World meteorological organization operational hydrology report no. 34. Hydrological models for water-resources system design and operation, https://library.wmo.int/doc_num.php?Explnum_id=1696 (last viewed 01.03.2022).

Bennett, N.D., Croke, B.F.W., Guariso, G., Guillaume, G.J.H.A., Hamilton, S.H, Jakeman, A.J., Marsili-Libelli, S., Newham, L.T.H, Norton, J.P., Perrin, C., Pierce, S.A., Robson, B., Seppelt, R., Voinov, A.A., Fath, B.D., Vazken Andreassian, V. 2013. Characterising performance of environmental models. Environmental Modelling & Software 40, 1-20 doi: 10.1016/j.envsoft.2012.09.011

Berendrecht, W. L., Heemink, A. W., Van Geer, F. C., Gehrels, J. C. 2003. Decoupling of modeling and measuring interval in groundwater time series analysis based on response characteristics. Journal of Hydrology 278(1-4), 1-16. doi: 10.1016/S0022-1694(03)00075-1

Bethere, L., Sennikovs, J., Bethers, U. 2017. Climate indices for the Baltic states from principal component analysis. Earth System Dynamics 8, 951-962. doi: 96210.5194/esd-8-951-2017

Beven, K. 1993. Prophecy, reality and uncertainty in distributed hydrological modelling. Advances in Water Resources 16(1), 41–51. doi:10.1016/0309-1708(93)90028-e

Bikše, J., Retiķe, I. 2018. An approach to delineate groundwater bodies at risk: Seawater intrusion in Liepāja (Latvia). E3S Web of Conferences. 54 00003, doi: 10.1051/e3sconf/20185400003

Brakkee, E., van Huijgevoort, M., Bartholomeus, R. P. 2021. Spatiotemporal development of the 2018–2019 groundwater drought in the Netherlands: a data-based approach. Hydrology and Earth System Sciences. Discussions. [preprint], in review, doi: 10.5194/hess-2021-64

Brunner, P., Simmons, C.T., Cook, P.G., Therrien, R. 2010. Modeling surface water-groundwater interaction with MODFLOW: Some Considerations. Groundwater 48, 174-180. doi: 10.1111/j.1745-6584.2009.00644.x

Bense, V. F., Gleeson, T., Loveless, S. E., Bour, O., Scibek, J. 2013. Fault zone hydrogeology. Earth-Science Reviews 127, 171-192. doi: 10.1016/j.earscirev.2013.09.008

Boschetto, R. G., Mohamed, R. M., Arrigotti, J. 2010. Vulnerability to desertification in a sub-Saharan region: a first local assessment in five villages of southern region of Malawi. Italian Journal of Agronomy 5(S3), 91-102. doi: 10.4081/ija.2010.s3.91

Cartwright, I., Cendón, D., Currell, M., Meredith, K. 2017. A review of radioactive isotopes and other residence time tracers in understanding groundwater recharge: Possibilities, challenges, and limitations. Journal of Hydrology, 555, 797-811. doi: 10.1016/j.jhydrol.2017.10.053

Collenteur, R.A. 2021a. How good is your model fit? Weighted goodness-of-fit metrics for irregular time series. Groundwater 59, 474-478. doi: 10.1111/gwat.13111

Collenteur, R.A. 2021b. Snow model, https://pastas.readthedocs.io/en/latest/examples/019_snowmodel.ipynb.html (last viewed: 16.03.2022).

Collenteur, R. A., Bakker, M., Caljé, R., Klop, S. A., Schaars, F. 2019. Pastas: Open source software for the analysis of groundwater time series. Ground water 57(6), 877–885. doi: 10.1111/gwat.12925

Collenteur, R., Bakker, M., Klammler, G., Birk, S. 2021. Estimating groundwater recharge from groundwater levels using non-linear transfer function noise models and comparison to lysimeter data. Hydrology and Earth System Sciences 25(5), 2931-2949. doi: 10.5194/hess-25-2931-2021

Condon, L. E., Maxwell, R. M. 2015. Evaluating the relationship between topography and groundwater using outputs from a continental-scale integrated hydrology model. Water Resources Research 51, 6602–6621. doi:10.1002/2014WR016774

Cornes, R., van der Schrier, G., van den Besselaar, E.J.M., Jones, P.D. 2018. An ensemble version of the E-OBS temperature and precipitation datasets. Journal of the Geophysical Research: Atmosphere, 123. doi:10.1029/2017JD028200

Cornu, J.F., Eme, D., Malard, F. 2013. The distribution of groundwater habitats in Europe. Hydrogeology Journal 21, 949–960. doi: 10.1007/s10040-013-0984-1

Devia, G. K., Ganasri, B. P., Dwarakish, G. S. 2015. A review on hydrological models. Aquatic procedia 4, 1001-1007. doi: 10.1016/j.aqpro.2015.02.126

Dēliņa, A. 2018. Pazemes ūdeņu resursi un ieguve (Groundwater resources and abstraction, in Latvian.) In Latvija. Zeme, daba, tauta, valsts [Latvia. Land, Nature, People, Country] (eds. Nikodemus, O., Kļaviņš, M., Krišjāne, Z., Zelčs, V.). ISBN: 9934182976

Duscher, K., Günther, A., Richts, A., Clos, P., Philipp, U., Struckmeier, W. 2015. The GIS layers of the "International Hydrogeological Map of Europe 1:1,500,000" in a vector format. Hydrogeology Journal 23, 1867–1875. doi: 10.1007/s10040-015-1296-4

Eriksson, M., Ebert, K., Jarsjö, J. 2018. Well salinization risk and effects of Baltic Sea level rise on the groundwater-dependent island of Öland, Sweden. Water 10(2), 141. doi: 10.3390/w10020141

Estonian Environmental Agency. 2022. Groundwater monitoring. History, https://keskkonnaagentuur.ee/keskkonnaagentuuri-tegevusvaldkonnad/vesi/pohjavesi#phjavee-seire (last viewed: 14.03.2022).

European Climate Assessment & Dataset. 2022. Countries, https://www.ecad.eu/countries/country_overview.php (last viewed 25.04.2022).

European Environmental Agency. 2018. European waters. Assessment of status and pressures 2018. EEA Report No 7/2018, https://www.eea.europa.eu/publications/state-of-water. (last viewed: 10.05.2021).

European Environmental Agency. 2022. CLC 2018, https://land.copernicus.eu/pan-european/corine-land-cover/clc2018?tab=metadata (last viewed: 26.02.2022).

Farthing, M.W., Ogden, F.L. 2017. Numerical solution of Richards' equation: A review of advances and challenges. Soil Science Society of America Journal 81, 1257-1269. doi: 10.2136/sssaj2017.02.0058

Fitts, C.R. 2002. Groundwater Science (Second Edition). Academic Press. ISBN 9780123847058

Forster, C., Smith, L, 1988. Groundwater flow systems in mountainous terrain: 2. Controlling factors. Water Resources Research 24(7), 1011–1023. doi:10.1029/wr024i007p01011

Freeze, R.A., Cherry, J.A. 1979. Groundwater. Prentice-Hall International. ISBN 0-13-365312-9

Gassman, P. W., Reyes, M. R., Green, C. H., Arnold, J. G. 2007. The soil and water assessment tool: historical development, applications, and future research directions. Transactions of the ASABE 50(4), 1211-1250. doi: 10.13031/2013.23637

Giese, M., Haaf, E., Heudorfer, B., Barthel, R. 2020. Comparative hydrogeology – reference analysis of groundwater dynamics from neighbouring observation wells. Hydrological Sciences Journal 65(10), 1685-1706. doi: 10.1080/02626667.2020.1762888

Gosk, E., Levins, I., Flindt Jørgensen, L. 2007. Shallow groundwater quality in Latvia and Denmark. GEUS Bulletin 13, 65–68, doi: 10.34194/geusb.v13.4979

Griebler, C., Avramov, M. 2015. Groundwater ecosystem services: a review. Freshwater Science 34(1), 355-367. doi: 10.1086/679903

Gupta, H., Kling, H. 2011. On typical range, sensitivity, and normalization of Mean Squared Error and Nash-Sutcliffe Efficiency type metrics. Water Resources Research 47, W10601. doi 10.1029/2011WR010962.

Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F. 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. Journal of Hydrology 377, (1–2), 80-91. doi: 10.1016/j.jhydrol.2009.08.003

Haaf, E., Giese, M., Heudorfer, B., Stahl, K., Barthel, R. 2020. Physiographic and climatic controls on regional groundwater dynamics. Water Resources Research 56, e2019WR026545. doi: 10.1029/2019WR026545

Hantush, M.S. 1956. Analysis of data from pumping tests in leaky aquifers. Transactions, American Geophysical Union 37(6), 702. doi:10.1029/TR037i006p00702

Hargreaves, G., Samani, Z. 1985. Reference crop evapotranspiration from temperature. Applied Engineering in Agriculture 1(2), 96-99. doi: 10.13031/2013.26773

Hariharan, V., Uma Shankar, M. 2017. A review of visual MODFLOW applications in groundwater modelling. IOP Conference Series: Materials Science and Engineering 263, 032025. doi: 10.1088/1757-899X/263/3/032025

Heudorfer, B., Haaf, E., Stahl, K., Barthel, R. 2019. Index-based characterization and quantification of groundwater dynamics. Water Resources Research 55, 5575–5592. doi: 10.1029/2018WR024418

Hokanson, K. J., Mendoza, C. A., Devito, K. J. 2019. Interactions between regional climate, surficial geology, and topography: Characterizing shallow groundwater systems in subhumid, low-relief landscapes. Water Resources Research 55, 284–297. doi: 10.1029/2018WR023934

Holman, I., Palmer, R., Leonavičiūtė, N. 2000. Using soil and Quaternary geological information to assess the intrinsic groundwater vulnerability of shallow aquifers: an example from Lithuania. Hydrogeology Journal 8, 636–645. doi: 10.1007/s100400000100

Hotta, N., Tanaka, N., Sawano, S., Kuraji, K., Shiraki, K., Suzuki, M. 2010. Changes in groundwater level dynamics after low-impact forest harvesting in steep, small watersheds. Journal of Hydrology, 385(1-4), 120-131. doi: 10.1016/j.jhydrol.2010.02.008

Hughes, J. D., Langevin, C. D., Banta, E. R. 2017. Documentation for the MODFLOW 6 framework (No. 6-A57). US Geological Survey. doi: 10.3133/tm6A57

Jaagus, J., Briede, A., Rimkus. E., Remm, K. 2010. Precipitation pattern in the Baltic countries under the influence of large-scale atmospheric circulation and local landscape factors. International Journal of Climatology 30(5), 708-720. doi:10.1002/joc.1929

Jackson, C. R., Wang, L., Pachocka, M., Mackay, J. D., Bloomfield, J. P. 2016. Reconstruction of multi-decadal groundwater level time-series using a lumped conceptual model. Hydrological Processes 30, 3107– 3125. doi: 10.1002/hyp.10850

Jajarmizadeh, M. 2012. A review on theoretical consideration and types of models in hydrology. Journal of Environmental Science and Technology 5, 249-261. doi: 10.3923/jest.2012.249.261

Jasiewicz, J. 2022. Calculate convergence index, https://grass.osgeo.org/grass78/ manuals/addons/r.convergence.html (last viewed: 03.04.2022).

Juodkazis, V. 1994. Groundwater quality and its monitoring in the Baltic States. GeoJournal 33(1), 63–70. doi: 10.1007/bf00810137

Kaleris, V. 1998. Quantifying the exchange rate between groundwater and small streams. Journal of Hydraulic Research 36(6), 913-932. doi: 10.1080/00221689809498593

Karofeld, E., Jarašius, L., Priede, A., Sendžikaitė, J. 2017. On the after-use and restoration of abandoned extracted peatlands in the Baltic countries. Restoration Ecology, 25(2), 293–300. doi:10.1111/rec.12436

Karro, E., Marandi, A., Vaikmäe, R., Uppin, M. 2009. Chemical peculiarities of the Silurian - Ordovician and Cambrian - Vendian aquifer systems in Estonia: an overview of hydrochemical studies. Estonian Journal of Earth Sciences 58(4), 342-352. doi: 10.3176/earth.2009.4.12

Kavetski, D., Kuczera, G. 2007. Model smoothing strategies to remove microscale discontinuities and spurious secondary optima in objective functions in hydrological calibration. Water Resources Research 43(3), W03411. doi: 10.1029/2006WR005195

Kitterød, N.O., Kvarner, J., Aagaard, P., Arustiene, J., de Beer, H., Bikše, J., Dagestad, A., Gundersen, P., Hansen, B., Hjartarson, A, Karro, E., Klavins, M., Marandi, A., Putyis, P., Radiene, R., Retike, I.,

Rossi, P., Thorling, L. Hydrologeology and Graundwater Quality in the Nordic Region. Hydrology Research Special Issue Paper OA, preprint.

Klints, I. Dēliņa, A. 2012. Groundwater abstraction in the Baltic Artesian Basin. In Highlights of groundwater research in the Baltic Artesian Basin (eds Dēliņa, A., Kalvāns, A., Saks, T., Bethers, U., Vircavs, V.). University of Latvia, pp.155.

Kløve, B., Allan, A., Bertrand, G., Druzynska, E., Ertürk, A., Goldscheider, N., Henry, S., Karakaya, N., Karjalainen, T., Koundouri, P., Knupfersberger, H., Kœvrner, K., Lundberg, A., Muotka, T., Preda, E., Pulido, Velazquez, M., Schipper, P. 2011. Groundwater dependent ecosystems. Part II. Ecosystem services and management in Europe under risk of climate change and land use intensification. Environmental Science & Policy, 14(7), 782-793. doi: 10.1016/j.envsci.2011.04.005.

Kottek, M., Grieser, J., Beck, C., Rudolf, B., Rubel, F. 2006. World Map of the Köppen-Geiger climate classification updated. Meteorologische Zeitschrift 15 (3), 259-263 doi: 10.1127/0941-2948/2006/0130

Knapp, C. L., Stoffel, T. L., Whitaker, S. D. 1980. Insulation solar radiation manual, Solar Energy Research Institute, Golden, Colo, https://www.nrel.gov/docs/legosti/old/3880.pdf (last viewed: 11.04.2022).

Knoben, W.J.M., Freer, J.E., Woods, R.A. 2019. Technical note: Inherent benchmark or not? Comparing Nash-Sutcliffe and Kling-Gupta efficiency scores. Hydrology and Earth System Sciences 23 (10), 4323—4331. 10.5194/hess-23-4323-2019

Krause, P., Boyle, D. P., Bäse, F. 2005. Comparison of different efficiency criteria for hydrological model assessment. Advances in Geosciences 5, 89-97. doi: 10.5194/adgeo-5-89-2005

Kriauciuviene, J., Meilutyte-Barauskiene, D., Reihan, A., Koltsova, T., Lita Lizuma, L., Sarauskiene, D. 2012 Variability in temperature, precipitation and river discharge in the Baltic States. Boreal Environment Research 17, 150–162. ISSN 1797-2469

Kroes, J., van Dam, J., Supit, I., de Abelleyra, D., Verón, S., de Wit, A., Boogaard, H., Angelini, M., Damiano F., Groenendijk, P., Wesseling, J., Veldhuizen, A. 2019. Agrohydrological analysis of groundwater recharge and land use changes in the Pampas of Argentina. Agricultural Water Management 213, 843-857. doi: 10.1016/j.agwat.2018.12.008

Krūmiņš, J., Kļaviņš, M., Dēliņa, A., Damkevics, R., Segliņš, V. 2021. Potential of the Middle Cambrian Aquifer for Carbon Dioxide Storage in the Baltic States. Energies 14(12), 3681. doi: 10.3390/en14123681

Kumar, N., Khamzina, A., Knöfel, P., Lamers, J., Tischbein, B. 2021. Afforestation of degraded croplands as a water-saving option in irrigated region of the Aral Sea basin. Water 13(10), 1433. doi: 10.3390/w13101433

Küttim, M., Küttim, L., Pajula, R. 2018. The current state and ecological restoration of peatlands in Estonia. Dynamiques environnementales. Journal international de géosciences et de l'environnement (42), 342-349. doi: 10.4000/dynenviron.2425

Langevin, C. D., Hughes, J. D., Banta, E. R., Niswonger, R. G., Panday, S., Provost, A. M. 2017. Documentation for the MODFLOW 6 groundwater flow model (No. 6-A55). US Geological Survey. doi: 10.3133/tm6A55

Latvian Geospatial Information Agency. S.a. Digital relief model, https://www.lgia.gov.lv/lv/Digit%C4%81lais%20reljefa%20modelis (last viewed: 14.03.2022).

Legates, D. R., McCabe, G. J. 1999. Evaluating the use of "goodness-of-fit" Measures in hydrologic and hydroclimatic model validation. Water Resources Research 35(1), 233–241. doi:10.1029/1998WR900018.

Lerner, D.N., Harris, B., 2009. The relationship between land use and groundwater resources and quality. Land Use Policy 26, S265-S273. doi: 10.1016/j.landusepol.2009.09.005

Levins, I., Levina, N., Gavena, I. 1998. Latvijas pazemes ūdeņu resursi (eng: Latvian Groundwater Resources). Valsts ģeoloģijas dienests, Rīga. ISBN 9984-9130-4-X

Liu, H.L., Chen, X., Bao, A., Wang, L. 2007. Investigation of groundwater response to overland flow and topography using a coupled MIKE SHE/MIKE 11 modeling system for an arid watershed. Journal of Hydrology 347 (3-4), 448–459. doi:10.1016/j.jhydrol.2007.09.053

Livada, I., Asimakopoulos, D. N. 2005. Individual seasonality index of rainfall regimes in Greece. Climate Research 28(2), 155–161. doi: 10.3354/cr028155

Lorenzo-Lacruz, J., Garcia, C., Morán-Tejeda, E. 2017. Groundwater level responses to precipitation variability in Mediterranean insular aquifers. Journal of Hydrology 552, 516-531. doi:10.1016/j.jhydrol.2017.07.011

Luoma, S., Okkonen, J. 2014. Impacts of future climate change and Baltic Sea level rise on groundwater recharge, groundwater levels, and surface leakage in the Hanko aquifer in southern Finland. Water 6(12), 3671-3700. doi: 10.3390/w6123671

Mackay, J.D., Jackson, C.R., Brookshaw, A., Scaife, A.A., Cook, J., Ward, R.S., 2015. Seasonal forecasting of groundwater levels in principal aquifers of the United Kingdom. Journal of Hydrology 530, 815-828. doi: 10.1016/j.jhydrol.2015.10.018

Mackay, J.D., Jackson, C.R., Wang, L. 2014. A lumped conceptual model to simulate groundwater level time-series. Environmental Modelling & Software 61, 229-245. doi: 10.1016/j.envsoft.2014.06.003

Makridakis, S., Hibon, M. 1997. ARMA models and the Box–Jenkins methodology. Journal of forecasting 16(3), 147-163. doi: 10.1002/(SICI)1099-131X(199705)16:3<147::AID-FOR652>3.0.CO;2-X

Marchant, B.P., Bloomfield, J.P. 2018. Spatio-temporal modelling of the status of groundwater droughts. Journal of Hydrology 564, 397-413. doi: 10.1016/j.jhydrol.2018.07.009

McDonald, M.G., Harbaugh, A.W. 2003. The History of MODFLOW. Groundwater 41, 280-283. doi: 10.1111/j.1745-6584.2003.tb02591.x

Melaku, N.D., Junye Wang, J. 2019. A modified SWAT module for estimating groundwater table at Lethbridge and Barons, Alberta, Canada. Journal of Hydrology 575, 420-431. doi: 10.1016/j.jhydrol.2019.05.052

Mokrik, R., Mažeika, J., Baublytė, A., Martma, T. 2009. The groundwater age in the Middle-Upper Devonian aquifer system, Lithuania. Hydrogeology Journal 17(4), 871-889. doi: 10.1007/s10040-008-0403-1

Moriasi, D.N, Gitau, M.W., Pai, N., Daggupati, P. 2015. Performance measures and evaluation criteria. Transactions of the ASABE 58(6), 1763-1785. doi: 10.13031/trans.58.10715

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., Veith, T. L. 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transactions of the ASABE 50(3), 885-900. doi: 10.13031/2013.23153

Nash, J.E., Sutcliffe, J.V. 1970. River flow forecasting through conceptual models. Part I: a discussion of principles. Journal of Hydrology 10, 282–290. doi: 10.1016/0022-1694(70)90255-6

Nistor, M.M. 2020. Groundwater vulnerability in Europe under climate change. Quaternary International 547, 185-196. doi: 10.1016/j.quaint.2019.04.012

Nygren, M., Giese, M., Kløve, B., Haaf, E., Rossi, P. M., Barthel, R. 2020. Changes in seasonality of groundwater level fluctuations in a temperate-cold climate transition zone. Journal of Hydrology X, 8, 100062. doi: 10.1016/j.hydroa.2020.100062

Obergfell, C., Bakker, M., Maas, K. 2019. Identification and explanation of a change in the groundwater regime using time series analysis. Ground Water 57(6), 886-894. doi: 10.1111/gwat.12891

O'Donnell, M.S., and Ignizio, D.A. 2012. Bioclimatic predictors for supporting ecological applications in the conterminous United States. US geological survey data series, 691(10), 4-9.

Open Street Map. 2022. Open Street Map Data Extracts, https://download.geofabrik.de/ (last viewed: 16.03.2022).

Orth, R., Staudinger, M., Seneviratne, S.I., Seibert, J., Zappa, M. 2015. Does model performance improve with complexity? A case study with three hydrological models. Journal of Hydrology 523, 147-159. doi: 10.1016/j.jhydrol.2015.01.044

Pakalne, M., Etzold, J., Ilomets, M., Jarašius, L., Pawlaczyk, P., Bociąg, K., Chlost, I., Cieśliński, R., Gos, K., Libauers, K., Pajula, R., Purre, A.H., Sendzikaite, J., Strazdina, L., Truus, L., Zableckis, N., Jurema, L., Kirschey, T. 2021. Best practice book for peatland restoration and climate change mitigation. Experiences from LIFE peat restore project. University of Latvia, Riga. ISBN: 987-9934-556-73-9

Paul, P. K., Zhang, Y., Ma, N., Mishra, A., Panigrahy, N., Singh, R. 2021. Selecting hydrological models for developing countries: Perspective of global, continental, and country scale models over catchment scale models. Journal of Hydrology 600, 126561. https://doi.org/10.1016/j.jhydrol.2021.126561

Peterson, T.J., Western, A.W., Cheng, X. 2018. The good, the bad and the outliers: automated detection of errors and outliers from groundwater hydrographs. Hydrogeology Journal 26(2), 371-280. doi: 10.1007/s10040-017-1660-7

Perens, R., Vallner, L. 1997. Water-bearing formation. In geology and mineral resources of Estonia (eds Raukas, A., Teedumae, A.). Estonian Academy Publishers, Tallinn. ISBN: 9985501853

Pezij, M., Augustijn, D.C.M., Hendriks, D.M.D., Hulscher, S.J.M.H. 2020. Applying transfer function-noise modelling to characterize soil moisture dynamics: a data-driven approach using remote sensing data. Environmental Modelling & Software 131, 104756. doi: 10.1016/j.envsoft.2020.104756

Pietrzak, D. 2021. Modeling migration of organic pollutants in groundwater — Review of available software. Environmental Modelling & Software 144, 105145. doi: 10.1016/j.envsoft.2021.105145

Pogumirskis, M., Sīle, T., Seņņikovs, J., Bethers, U. 2021. PCA analysis of wind direction climate in the Baltic states. Tellus A: Dynamic Meteorology and Oceanography 73(1), 1-16. doi: 10.1080/16000870.2021.1962490

Post, V., Kooi, H., Simmons, C. 2007. Using hydraulic head measurements in variable-density ground water flow analyses. Groundwater 45, 664-671. doi: 10.1111/j.1745-6584.2007.00339.x

Post, V.E., and J.R. von Asmuth. 2013. Hydraulic head measurements - New technologies, classic pitfalls. Hydrogeology Journal 21 (4), 737–750. doi: /10.1007/s10040-013-0969-0

Ptak, T., Piepenbrink, M., Martac, E. 2004. Tracer tests for the investigation of heterogeneous porous media and stochastic modelling of flow and transport—a review of some recent developments. Journal of hydrology 294(1-3), 122-163. doi: 10.1016/j.jhydrol.2004.01.020.

Raidla, V., Pärn, J., Aeschbach, W., Czuppon, G., Ivask, J., Kiisk, M., et al. 2019. Intrusion of saline water into a coastal aquifer containing palaeogroundwater in the Viimsi peninsula in Estonia. Geosciences 9(1), 47. doi: 10.3390/geosciences9010047

Rasmussen, T.C., Crawford, L.A. 1997. Identifying and removing barometric pressure effects in confined and unconfined aquifers. Groundwater 35, 502-511. doi: 10.1111/j.1745-6584.1997.tb00111.x

Refsgaard, J. C., Knudsen, J. 1996. Operational validation and intercomparison of different types of hydrological models. Water Resources Research 32 (7), 2189– 2202. doi:10.1029/96WR00896

Republic of Estonia Land Board. 2022. Elevation data, https://geoportaal.maaamet.ee/eng/Spatial-Data/Elevation-data-p308.html (last viewed: 14.03.2022).

Retike, I., Bikše, J., Kalvāns, A., Dēliņa, A., Avotniece, Z., Zaadnoordijk, W. J., Jemeļjanova, M., Zelenkevičs, A., Baikovs, A. 2022. Rescue of groundwater level time series: How to visually identify and treat errors. Journal of Hydrology 605, 127294. doi: 10.1016/j.jhydrol.2021.127294

Rinderer, M., van Meerveld, I., Stähli, M., Seibert, J. 2016. Is groundwater response timing in a pre-alpine catchment controlled more by topography or by rainfall? Hydrological Processes 30(7), 1036–1051. doi:10.1002/hyp.10634

Samani, Z. 2000. Estimating solar radiation and evapotranspiration using minimum climatological data. Journal of Irrigation and Drainage Engineering 126(4), 265-267. doi:10.1061/(ASCE)0733-9437(2000)126:4(265)

Schaefli, B., Gupta, H. V. 2007. Do Nash values have value? Hydrological Processes 21, 2075-2080. doi: 10.1002/hyp.6825

Stekhoven, D.J., Buehlmann, P. 2012. MissForest - nonparametric missing value imputation for mixed-type data. Bioinformatics 28(1) 2012, 112-118. doi: 10.1093/bioinformatics/btr597

Sharma, S.K., Gajbhiye, S. Tignath, S. 2015. Application of principal component analysis in grouping geomorphic parameters of a watershed for hydrological modeling. Applied Water Science 5, 89–96 doi: 10.1007/s13201-014-0170-1

Silberstein, R.P. 2006. Hydrological models are so good, do we still need data? Environmental Modelling & Software 21 (9), 1340-1352. doi: 10.1016/j.envsoft.2005.04.019

Simunek, J., Jacques, D., Langergraber, G., Bradford, S. A., Šejna, M., van Genuchten, M. T. 2013. Numerical modeling of contaminant transport using HYDRUS and its specialized modules. Journal of the Indian institute of science 93(2), 265-284. ISSN: 0970-4140 Coden-JIISAD

Simunek, J., Van Genuchten, M. T., Šejna, M. 2012. HYDRUS: Model use, calibration, and validation. Transactions of the ASABE 55(4), 1263-1274. doi: 10.13031/2013.42239

Singh, V. P. 2018. Hydrologic modeling: progress and future directions. Geoscience letters 5(1), 1-18. doi: 10.1186/s40562-018-0113-z

Sliaupa, S., Shogenova, A., Shogenov, K., Sliaupiene, R., Zabele, A., Vaher, R. 2008. Industrial carbon dioxide emissions and potential geological sinks in the Baltic states. Oil shale, 25(4). doi: 10.3176/oil.2008.4.06

Smerdon, B. D., Redding, T., Beckers, J. 2009. An overview of the effects of forest management on groundwater hydrology. Journal of Ecosystems and Management 10(1), 22-44. doi: 10.22230/jem.2009v10n1a409

Soleimani Motlagh, M., Ghasemieh, H., Talebi, A., Abdollahi, K. 2017. Identification and analysis of drought propagation of groundwater during past and future periods. Water resources management 31(1), 109-125. doi: 10.1007/s11269-016-1513-5

Solomatine, D.P., Wagener T. 2011. Hydrological modeling. In Treatise on Water Science (eds Wilderer, P.). Elsevier Science, ISBN 978-0-444-53199-5.

Sorensen, R., Zinko, U., Seibert, J. 2006. On the calculation of the topographic wetness index: evaluation of different methods based on field observations. Hydrology and Earth System Sciences Discussions 2(4), 1807-1834. doi: 10.5194/hess-10-101-2006

Sutanto, S. J., Van Lanen, H. A. 2021. Streamflow drought: implication of drought definitions and its application for drought forecasting. Hydrology and Earth System Sciences 25(7), 3991-4023. doi: 10.5194/hess-25-3991-2021

Vaikmäe, R., Vallner, L., Loosli, H.H., Blaser, P.C., Juillard-Tardent, M. 2001. Palaeogroundwater of glacial origin in the Cambrian-Vendian aquifer of northern Estonia. Geological Society, London, Special Publications 189(1), 17–27. doi:10.1144/GSL.SP.2001.189.01.03

Van Loon, A.F. 2015. Hydrological drought explained. Wiley Interdisciplinary Reviews: Water 2(4), 359-392. doi: 10.1002/wat2.1085

Van Loon, A. F., Van Lanen, H. A., Hisdal, H. E. G. E., Tallaksen, L. M., Fendeková, M., Oosterwijk, J., Horvat, O., Machlica, A. 2010. Understanding hydrological winter drought in Europe. Global Change: Facing Risks and Threats to Water Resources. IAHS Publication, 340, 189-197.

von Asmuth, J. R. 2012. Groundwater system identification through time series analysis. Ph.D. thesis. Delft University of Technology. ISBN 978-90-5155-079-5

von Asmuth, J. R., Bierkens, M.F.P., K. Maas, K. 2002. Transfer function-noise modeling in continuous time using predefined impulse response functions. Water Resources Research 38(12), 1287. doi: 10.1029/2001WR001136

von Asmuth, J.R., Maas, K., Bakker, M., Petersen, J. 2008. Modeling time series of ground water head fluctuations subjected to multiple stresses. Ground Water 46(1), 30-40. doi: 10.1111/j.1745-6584.2007.00382.x.

Wada, Y., van Beek, L.P.H., van Kempen, C.M., Reckman, J.W.T.M., Vasak, S., Bierkens, M.F.P. 2010. Global depletion of groundwater resources. Geophysical research letters 37(20). doi: 10.1029/2010GL044571

Wang, Y., Jiang, R., Xie, J., Zhao, Y., Yan, D., Yang, S., 2019. Soil and water assessment tool (SWAT) model: A systemic review. Journal of Coastal Research 93(SI), 22–30. doi: 10.2112/SI93-004.1

Wendland, F., Blum, A., Coetsiers, Gorova, M. R., Griffioen, J., Grima J., Hinsby K., Kunkel R., Marandi, A., Melo, T., Panagopoulos, A., Pauwels, H., Ruisi, M., Traversa, P., Vermooten, P.J.A., Walraevens, K.. 2008. European aquifer typology: a practical framework for an overview of major groundwater composition at European scale. Environmental Geology 55, 77–85. doi: 10.1007/s00254-007-0966-5

Yevjevich., V. 1987. Stochastic models in hydrology. Hydrology and Hydraulics 1(1), 17–36. doi:10.1007/bf01543907

Yuan, Y., Koropeckyj-Cox, L. 2022. SWAT model application for evaluating agricultural conservation practice effectiveness in reducing phosphorous loss from the Western Lake Erie Basin. Journal of Environmental Management 302, 114000. doi: 10.1016/j.jenvman.2021.114000

Zaadnoordijk, W.J., Bus, S.A., Lourens, A., Berendrecht, W.L. 2019. Automated time series modeling for piezometers in the national database of the netherlands. Groundwater 57, 834-843. doi: 10.1111/gwat.12819

Zaidman, M. D., Rees, H. G., Young, A. R. 2002. Spatio-temporal development of streamflow droughts in north-west Europe. Hydrological Earth System Sciences 6, 733–751. doi: 10.5194/hess-6-733-2002

Zeiger, S.J., Owen, M.R., Pavlowsky, R.T. 2021. Simulating nonpoint source pollutant loading in a karst basin: A swat modeling application. Science of The Total Environment, 785, 147295. doi: 10.1016/j.scitotenv.2021.147295

Zektser, I.S., Loaiciga, H.A. 1993. Groundwater fluxes in the global hydrologic cycle: past, present and future. Journal of Hydrology 144(1-4), 405–427. doi:10.1016/0022-1694(93)90182-9

Zhang, H., Hiscock, K.M. 2010. Modelling the impact of forest cover on groundwater resources: A case study of the Sherwood Sandstone aquifer in the East Midlands, UK. Journal of Hydrology 392(3-4), 136–149. doi:10.1016/j.jhydrol.2010.08.002

Annex 1. The categories and assigned values of hydrograph descriptors

Typology component	Typology subcomponent	Assigned values	Comments
Structure	Seasonality magnitude - how prominent is the seasonal influence	Strong	Seasonality visible in all years
		Moderate/ irregular	Seasonality is visible in some/ most years
		Minimal to none	No visible seasonal variability
	Flashiness – occurrence of peaks (dominant peaks)	High	Changes more than once per month (jagged pattern)
		Moderate	Yearly peaks and troughs with few exceptions
		Low	Changes in less than a year
	Flashiness – occurrence of peaks (secondary peaks)	High	Daily to multi-day peaks.
		Moderate	Weekly to monthly peaks
		Low	No or next to no peaks.
	Inter-annual variability – how prominent are multi-year patterns	strong	IAV that has a strong impact on the level variability
		Moderate	Minor trends or inter-annual fluctuations
		Minimal to none	Stationery series
Distribution	Boundedness-tendency towards either extreme	[0; 1]	Median of the normalized series
	Density-dispersion around modes	[0; 1]	Inter-quartile range (IQR) of the normalized series
	Modality	[0; 1]	Mode of the normalized series

Annex 2. R^2 values of linearly regressed NSE values to environmental descriptors

	LG	L4	RZE	RZS
descriptor				
IQR	0.066***	0.1***	0.085***	0.086***
PE_avg	0.005	0.04***	0.04***	0.005
P_avg	0.01*	0.021**	0.023**	0.019**
SI	0.004	0.0	0.001	0.002
T_avg	0.034***	0.103***	0.05***	0.004
agricult	0.002	0.01	0.023**	0.045***
amplitude	0.002	0.001	0.005	0.012*
aq_depth	0.083***	0.008	0.051***	0.025***
aq_thickness	0.01*	0.003	0.007	0.014*
area_bel_w	0.022**	0.0	0.009	0.001
artific	0.002	0.0	0.001	0.0
catch_flowacc	0.0	0.009	0.001	0.001
clay_marl_silt	0.059***	0.005	0.029***	0.012*
converg	0.005	0.004	0.005	0.001
depth_to_gw	0.058***	0.002	0.05***	0.037***
diff_well_bas	0.006	0.0	0.002	0.001
diff_well_riv	0.018**	0.016**	0.019**	0.025***
dist_bas	0.0	0.002	0.003	0.002
dist_riv	0.0	0.0	0.001	0.001
dol_limestone	0.015**	0.005	0.003	0.015**
forests	0.005	0.003	0.023**	0.03***
gw_level_mean	0.008	0.001	0.001	0.003
gypsum	0.0	0.001	0.0	0.002
igneous	0.001	0.001	0.001	0.001
median	0.0	0.032***	0.015**	0.037***
peat	0.037***	0.001	0.009	0.008
sandstone_siltstone	0.058***	0.013*	0.039***	0.009
silt_loam	0.011*	0.005	0.039***	0.022**

technogenic_soil	0.0	0.004	0.005	0.0
terrigenous	0.02**	0.01*	0.01*	0.006
twi	0.006	0.009	0.031***	0.021**
water	0.052***	0.047***	0.051***	0.063***
water_voids	0.002	0.001	0.0	0.004
well_h_range	0.0	0.004	0.0	0.0
wetlands	0.008	0.003	0.001	0.003

Annex 3. R^2 values of linearly regressed model parameter 'A' values to environmental descriptors

	LG	L4	RZE	RZS
descriptor				
IQR	0.02	0.011	0.056	0.026
PE_avg	0.058	0.008	0.001	0.036
P_avg	0.067	0.006	0.0	0.059
SI	0.006	0.024	0.157	0.022
T_avg	0.026	0.07	0.111	0.068
agricult	0.036	0.048	0.089	0.01
amplitude	0.2*	0.0	0.067	0.066
aq_depth	0.277**	0.123*	0.002	0.001
aq_thickness	0.092	0.0	0.002	0.014
area_bel_w	0.117	0.149*	0.073	0.005
artific	0.003	0.049	0.011	0.068
catch_flowacc	0.025	0.004	0.03	0.046
clay_marl_silt	0.013	0.113	0.029	0.037
converg	0.312**	0.033	0.021	0.131
depth_to_gw	0.003	0.216**	0.451***	0.069
diff_well_bas	0.069	0.047	0.01	0.053
diff_well_riv	0.023	0.008	0.026	0.153
dist_bas	0.192*	0.004	0.018	0.02
dist_riv	0.001	0.003	0.038	0.272**
dol_limestone	0.0	0.06	0.072	0.059
forests	0.069	0.0	0.101	0.003
gw_level_mean	0.054	0.001	0.052	0.006
gypsum	0.0	0.387***	-0.0	0.0
igneous	0.0	0.0	-0.0	0.0
median	0.02	0.005	0.037	0.065
peat	0.0	0.002	-0.0	0.024
sandstone_siltstone	0.445***	0.068	0.001	0.005
silt_loam	0.017	0.015	0.017	0.035

technogenic_soil	0.051	0.013	0.024	0.024
terrigenous	0.152	0.016	0.022	0.149
twi	0.02	0.007	0.093	0.067
water	0.05	0.0	0.039	0.014
water_voids	0.0	0.103	-0.0	0.0
well_h_range	0.118	0.094	0.073	0.216*
wetlands	0.029	0.018	0.0	0.0

Annex 4. R^2 values of linearly regressed model parameter 'n' values to environmental descriptors

	LG	L4
descriptor		
IQR	0.056	0.08
PE_avg	0.031	0.138*
P_avg	0.096	0.0
SI	0.003	0.003
T_avg	0.095	0.007
agricult	0.016	0.266***
amplitude	0.253**	0.057
aq_depth	0.316**	0.066
aq_thickness	0.071	0.093
area_bel_w	0.039	0.065
artific	0.024	0.051
catch_flowacc	0.044	0.021
clay_marl_silt	0.009	0.109
converg	0.307**	0.0
depth_to_gw	0.006	0.0
diff_well_bas	0.005	0.0
diff_well_riv	0.329**	0.032
dist_bas	0.238**	0.003
dist_riv	0.04	0.0
dol_limestone	0.003	0.156*
forests	0.005	0.025
gw_level_mean	0.025	0.01
gypsum	0.0	0.032
igneous	0.0	-0.0
median	0.048	0.162*
peat	0.0	0.01
sandstone_siltstone	0.283**	0.012
silt_loam	0.006	0.013

technogenic_soil	0.203*	0.11
terrigenous	0.254**	0.006
twi	0.08	0.001
water	0.062	0.0
water_voids	0.0	0.386***
well_h_range	0.025	0.007
wetlands	0.015	0.021

Annex 5. R^2 values of linearly regressed model parameter ‘a’ value to environmental descriptors

	LG	L4	RZE	RZS
descriptor				
IQR	0.056	0.0	0.02	0.111
PE_avg	0.031	0.003	0.058	0.041
P_avg	0.096	0.041	0.067	0.194*
SI	0.003	0.074	0.006	0.059
T_avg	0.095	0.005	0.026	0.182*
agricult	0.016	0.11	0.036	0.024
amplitude	0.253**	0.002	0.2*	0.022
aq_depth	0.316**	0.003	0.277**	0.004
aq_thickness	0.071	0.056	0.092	0.001
area_bel_w	0.039	0.0	0.117	0.005
artific	0.024	0.087	0.003	0.009
catch_flowacc	0.044	0.061	0.025	0.01
clay_marl_silt	0.009	0.001	0.013	0.394***
converg	0.307**	0.003	0.312**	0.276**
depth_to_gw	0.006	0.045	0.003	0.163
diff_well_bas	0.005	0.012	0.069	0.153
diff_well_riv	0.329**	0.012	0.023	0.021
dist_bas	0.238**	0.002	0.192*	0.209*
dist_riv	0.04	0.032	0.001	0.223*
dol_limestone	0.003	0.06	0.0	0.054
forests	0.005	0.001	0.069	0.027
gw_level_mean	0.025	0.038	0.054	0.269**
gypsum	0.0	0.032	0.0	0.033
igneous	0.0	0.0	0.0	0.0
median	0.048	0.222**	0.02	0.079
peat	0.0	0.022	0.0	0.312**
sandstone_siltstone	0.283**	0.013	0.445***	0.068
silt_loam	0.006	0.038	0.017	0.283**

technogenic_soil	0.203*	0.006	0.051	0.027
terrigenous	0.254**	0.009	0.152	0.295**
twi	0.08	0.001	0.02	0.096
water	0.062	0.002	0.05	0.054
water_voids	0.0	0.123*	0.0	0.033
well_h_range	0.025	0.014	0.118	0.513***
wetlands	0.015	0.016	0.029	0.0

Non-exclusive licence to reproduce thesis and make thesis public

I, Marta Jemeļjanova,

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright,

"Evaluating impulse response function models for groundwater level modeling across the Baltic states",

supervised by *Ph.D.* Alexander Kmoch, *MSc* Jānis Bikše, *MSc* Raoul Collenteur.

2. I grant the University of Tartu a permit to make the work specified in p. 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 3.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.

3. I am aware of the fact that the author retains the rights specified in p. 1 and 2.

4. I certify that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Marta Jemeļjanova

30/05/2022