# Efficient screening of groundwater head monitoring data for anthropogenic effects and measurement errors

5 Christian Lehr[1,2], Gunnar Lischeid[1,2]

[1] Leibniz Centre for Agricultural Landscape Research (ZALF), Müncheberg, Germany

[2] University of Potsdam, Institute of Environmental Science and Geography, Potsdam, Germany

Correspondence to: C. Lehr (lehr@zalf.de)

**Abstract**

10    Groundwater levels are monitored by environmental agencies to support sustainable use of groundwater resources. For this purpose continuous and spatially comprehensive monitoring in high spatial and temporal resolution is desired. This leads to large data sets that have to be quality checked and analysed to distinguish local anthropogenic influences from natural variability of the groundwater level dynamics at each well. Both technical problems with the measurements as well as local anthropogenic influences can lead to local anomalies in the hydrographs. We suggest a fast and efficient screening method

15    for identification of well-specific peculiarities in hydrographs of groundwater head monitoring networks. The only information required is a set of time series of groundwater head all measured at the same instants of time. For each well of the monitoring network a reference hydrograph is calculated, describing expected "normal" behaviour at the respective well as it is typical for the monitored region. The reference hydrograph is calculated by multiple linear regression of the observed hydrograph with the "stable" principal components (PCs) of a principal component analysis of all groundwater head series of

20    the network as predictor variables. The stable PCs are those PCs which were found in a random subsampling procedure to be rather insensitive to the specific selection of analysed observation wells, i.e. complete series, and to the specific selection of measurement dates. Hence they can be considered to be representative for the monitored region in the respective period. The residuals of the reference hydrograph describe local deviations from the normal behaviour. Peculiarities in the residuals allow the data to be checked for measurement errors, and the wells with a possible anthropogenic influence to be identified.

25    The approach was tested with 141 groundwater head time series from the state authority groundwater monitoring network in northeast Germany covering the period from 1993 to 2013 at an approximately weekly frequency of measurement.

**Short summary**

A screening method for fast identification of well-specific peculiarities in hydrographs of groundwater head monitoring

30    networks is suggested and tested. The only information required is a set of time series of groundwater head readings all measured at the same instants of time. The results were used to quality check the data for measurement errors and to identify wells with possible anthropogenic influence.

**Running title:**

35    Efficient screening of groundwater head monitoring data

**1 Introduction**

Sustainable management of groundwater resources aims to ensure that the abstraction of groundwater does not exceed groundwater recharge over the long-term (e.g., Gleeson et al., 2012). This resonates in the Water Framework Directive of the

40 European Union which aims to achieve and maintain a good status of groundwater quantity; it includes the obligation to monitor the temporal development of groundwater storage with sufficient spatial and temporal resolution to be able to distinguish between anthropogenic influences and natural variability (EU, 2000).

In practice declining trends in observed groundwater head time series serve often as first indication for anthropogenic effects. However, declining trends in hydrological systems do not necessarily imply anthropogenic influence, but might be

45 rather indications of naturally occurring long term persistence (Hurst, 1951; Mandelbrot and Ness, 1968; Mandelbrot and Wallis, 1969) induced for example by the natural fluctuation of climatic drivers (e.g., Koutsoyiannis, 2006). Thus continuity of hydrologic records for more than "just" a few decades is mandatory to incorporate this issue in water management (Hirsch, 2011). This holds the more for groundwater monitoring due to the much more pronounced filtering of short term fluctuations in the subsurface compared to precipitation, soil moisture or streamflow (Skøien et al., 2003).

50 Thus, sustainable management of groundwater resources requires a spatially differentiated and comprehensive monitoring of the groundwater level covering its temporal development continuously over decades which leads to large data sets. Checking the data quality at the wells of a spatially comprehensive network can be very time consuming. Measurement errors as well as local anthropogenic influences lead to anomalies at individual sites. Thus, local anomalies in the hydrographs can serve as indication for both aspects. That requires a reference, either in the form of observed hydrographs measured at specific

55 observation wells which are considered representative for undisturbed behaviour typical for the region (e.g., Winter et al., 2000; Gangopadhyay et al., 2001), or in form of some kind of modelled regional reference hydrograph. Here, we focus on the second case.

One option is to apply a physically-based model for that purpose. Depending on the characteristics of the region and the level of model complexity, the amount of data required can be large (cf. discussion in Coppola et al., 2003). A non-

60 exhaustive list of typical basic requirements includes climatological data to drive the model, data on the hydraulic properties of the subsurface which could comprise various aquifers, data on land use or time series of water abstraction. That information is required in a spatially distributed manner. In many cases, the monitoring effort, the effort to set up the model, the complexity of the model and the demand on data are substantial obstacles for environmental authorities or consultants at larger spatial scales.

65 Consequently, in practice some model parameters will serve during calibration as surrogate for missing data. For example, Wriedt (2017) calculated for each observation well in a monitoring network a "theoretical climatic hydrograph" by fitting the monthly climatic water balance to the observed groundwater hydrograph using a damping and a translation factor. Here, the fitting of the damping and the translation factor compensated for missing information on different properties of the analysed groundwater system like different substrates, flow paths, etc.

70 Another option is to fit empirical models based on the relationships between easy to obtain independent variables and the observed water level. This has been performed for example by means of multiple linear regression (Hodgson, 1978), artificial neural networks (Coulibaly et al., 2001; Coppola et al., 2003), the combination of an artificial neural network and a linear autoregressive model with exogenous input (Wunsch et al., 2018) or different other approaches from the field of artificial intelligence like adaptive neuro-fuzzy inference system, genetic programming, support vector machine or hybrid

75 models such as wavelet-artificial intelligence models (Rajaee et al., 2019). Others applied time series models like for example the predefined impulse response function in continuous time models (Von Asmuth et al., 2008), mixed models which combine deterministic fixed effects models with statistical random effects models (Marchant et al., 2016), or the combination of different methods from the fields of exploratory data analysis, information theory and machine learning (Sahoo et al., 2017). Those data-driven approaches make efficient use of the available data and are therefore recommended

as a relatively cost and time efficient way to model groundwater level in areas with limited data (Hodgson, 1978; Coulibaly et al., 2001; Coppola et al., 2003). Another benefit is that the respective models can easily be updated once new measurements or additional variables become available (Coppola et al., 2003).

Here, we suggest to derive local reference time series based on a Principal Component Analysis (PCA) of a set of monitored time series. This approach does not require any other data than the time series of water level all measured at the same instants of time. PCA is one of the most established, fastest and computationally cheapest statistical approaches to summarize the spatiotemporal variability of a set of spatially distributed time series. Based on the linear correlation structure of the data set linearly independent principal components (PCs) are derived. Each PC is associated with one characteristic spatiotemporal pattern. Except from being fast and computationally very cheap, PCA is readily implemented in most of the common statistical software packages.

In analogy to climatology (e.g. Richman, 1986), the spatiotemporal patterns of the PCs have been used for long in hydrology as a compact description of the dominant modes of hydrological variability in a region. Pioneering studies used it to describe dominant modes of streamflow variability in the European USSR (Smirnov, 1972, 1973), USA and Southern Canada (Bartlein, 1982), the USA alone (Lins, 1985a,b, 1997) and different regions of Sweden (Gottschalk, 1985). The leading PCs were used in combination with cluster analysis for the classification of streamflow (Hannah et al., 2000) and groundwater (Triki et al., 2014) hydrographs into groups with similar dynamics, which allowed for example to reduce the effort of modelling of all the time series in a groundwater monitoring network to a few representative hydrographs (Upton and Jackson, 2011). These approaches have to be distinguished from using PCs for the identification of prevailing processes or functional relationships (e.g., Longuevergne et al., 2007; Lewandowski et al., 2009; Hohenbrink et al., 2016) which is beyond the scope of this paper.

To our knowledge, the application of PCA based approaches in the context of compact description of hydrological variability focussed so far mainly on the leading PCs, that is, the main modes of hydrological variability on the scale of the analysed data set, like the main regional spatiotemporal patterns in a monitoring network. For example, Smirnov (1973) used the leading PCs for filtering small scale disturbances from the large scale patterns of long-period streamflow fluctuations in the European USSR. In groundwater monitoring, the leading PCs were used for example in the evaluation of groundwater monitoring networks to identify the observation wells which were most representative for the analysed region (Winter et al., 2000; Gangopadhyay et al., 2001).

In this study, we used the leading principal components (PCs) of a set of groundwater head time series of a large scale monitoring network to decompose the observed hydrographs into a reference part and a residual part for each site separately. The reference hydrograph of an observation well describes the expected "normal" behaviour at that well, i.e. the part of the observed hydrograph which is considered typical for the monitored region. It is determined by multiple linear regression of the observed hydrograph with the "stable" PCs as predictor variables. The "stable" PCs are determined by comparing the results of a series of PCAs which were performed with different randomly selected subsets of the complete data set. The sub data sets were derived in the spatial domain as random selections of the observation wells, i.e. complete series, and in the temporal domain as random selections of the measurement dates. Those PCs that turned out to be rather insensitive to the specific selection of analysed wells and measurement dates were defined as sufficiently stable and considered to be representative for the monitored region in the analysed period. The residual part of the hydrograph depicts the local deviations from the normal behaviour at the respective observation well, which is then analysed for peculiarities pointing to technical problems or anthropogenic effects. Other applications of the reference hydrograph, like gap-filling in series which were not included in the PCA, are shortly discussed. In contrast to other PCA applications this approach does neither require any interpretation of the leading PCs nor any explicit spatial analysis. The approach was tested with 141 groundwater head time series of the authority's groundwater monitoring network of the German Federal State of Mecklenburg-Vorpommern covering the period from 1993 to 2013 in approximately weekly resolution.

## 2 Data

### 2.1 Study region

The Federal State of Mecklenburg-Vorpommern is located in the north-east of Germany (Figure 1) covering an area of 23,214 km² (Statistikportal, 2018). The hydrogeological structure in the area consists of several "regional aquifer systems" of Pleistocene origin that are considered to be generally hydraulically separated (Manhenke et al., 2001). The term "regional aquifer system" describes a system of aquifers of different stratigraphy which are hydraulically connected (Manhenke et al., 2001). Those were formed during different periods of glaciation. Thus, the stratigraphy, or the sequence of stratigraphies, within one aquifer system is not everywhere the same (Manhenke et al., 2001). The water management in the region is based on the aquifer systems and not on single aquifers.

Single aquifers can be horizontally as well as vertically separated. The aquifers consist mainly of sandy and gravelly sediments, the intermediate aquitards mainly of till. The Pleistocene sediments usually overlie Tertiary sediments and can comprise more than 100 m. Within the Tertiary sediments the Rupelian clay layer hydraulically separates the underlying saline Tertiary groundwater from the fresh groundwater (Manhenke et al., 2001). However, at some locations upwelling salty groundwater reaches the surface (compare LUNG (1984) and figure 2.11-6 in Martens and Wichmann (2007)), indicating that the hydraulic separation of the regional aquifer systems is not everywhere complete.

Groundwater in Mecklenburg-Vorpommern is monitored by the federal state office of Environment, Nature Conservation and Geology (LUNG). Monitoring comprised the uppermost three regional aquifer systems, but not all of them are contiguous (Hennig et al., 2002). In addition to the regional aquifer systems, in some areas shallow local aquifers with an extent of usually a few km$^2$, occasionally more than 100 km$^2$, have been identified and are monitored as well (Hilgert and Hennig, 2017). These local aquifers are not strictly hydraulically decoupled from the regional aquifer systems, but their hydraulic connection is inhibited (Hilgert and Hennig, 2017). A mean annual groundwater recharge of +122.3 mm for the whole Mecklenburg-Vorpommern was estimated for the years 1971 to 2000 by Hilgert (2009) based on the work of Hennig and Hilgert (2007). A map of the spatial distribution of the mean annual groundwater recharge is provided by the LUNG (LUNG, 2009).

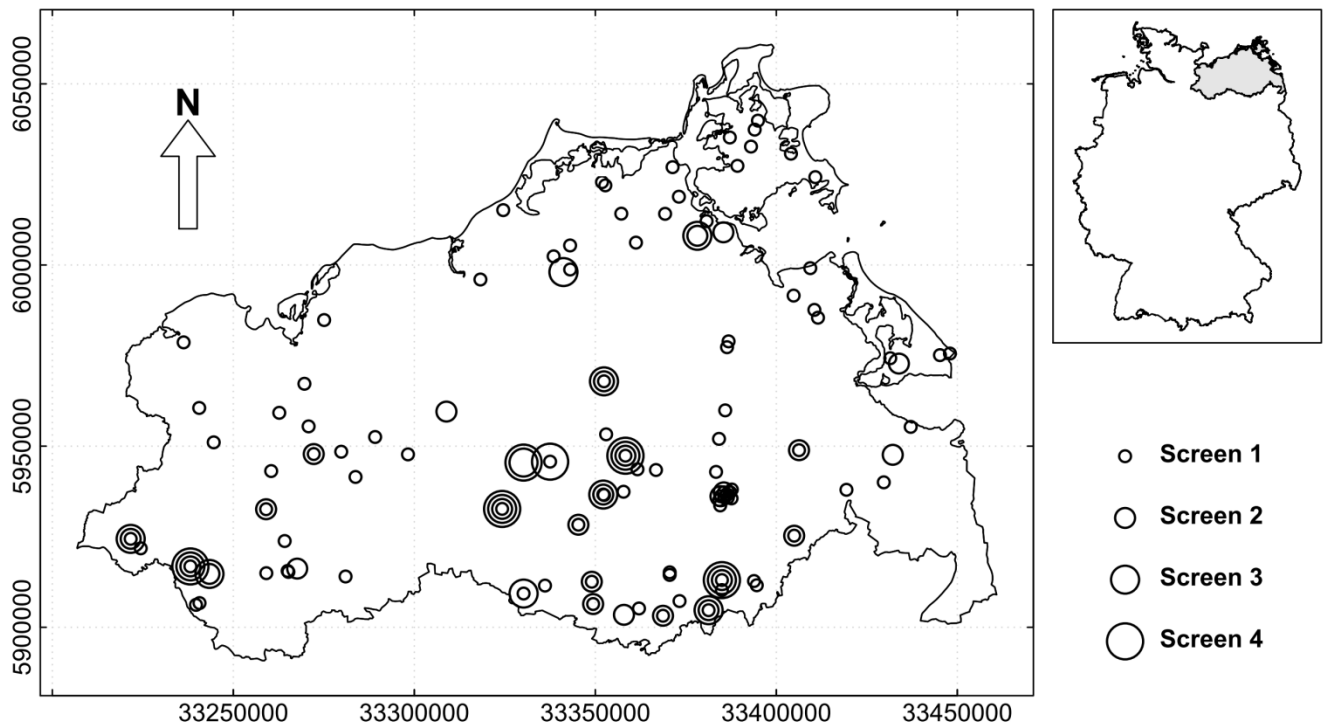



**Figure 1: Map of the study area and the selection of groundwater observation wells (N=141) in the federal state of Mecklenburg-Vorpommern (highlighted by grey shading in the inset map). The administrative borders of Germany and of the federal state of Mecklenburg-Vorpommern were obtained from GADM (2017).**

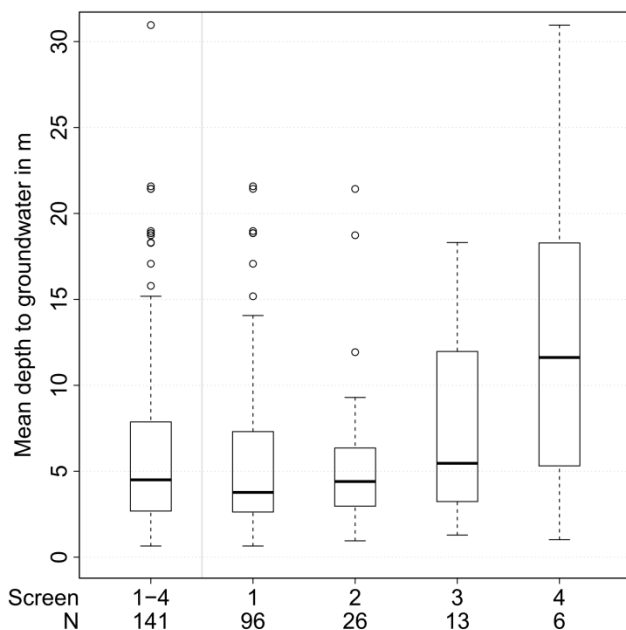
## 2.2 Groundwater head data and preprocessing

Groundwater level data were available for 583 observation wells in the monitoring network of the LUNG. In general, the data set covered a range of different monitoring periods. The mean length of measurement record was approximately 19.9 years. At some wells monitoring started as early as 1953.

For this study, we selected those 141 wells which were continuously monitored during the 20 years period from 1993-11-01 to 2013-10-22 (Figure 1). All 141 groundwater head series were quality checked before by the LUNG. The analysed data set comprised only those time series which were considered to be without any direct anthropogenic effects or measurement errors. During the monitoring period the state office aimed to take measurements consistently on the $1^{st}$, $8^{th}$, $15^{th}$ and $22^{nd}$ day of the month. Exceptions from this rule appeared for eight series (Figure S1). Bi-weekly readings were conducted from the beginning of the monitoring period at two wells till November 2005 and at one well till December 2006. Monthly readings were conducted from the beginning of the monitoring period at two wells till November 2007, at one well till September 2001 and at another well till December 2003. At one well the reading interval was changed to monthly readings in May 2002 and continued so till the end of the monitoring period. The variety of days between subsequent readings is shown for each series in Figure S2. The mean, minimum and maximum of all 141 series' mean measurement intervals were 5.3, 1.5 and 13.6 days. The mean, minimum and maximum of all 141 series' maximum time gaps between subsequent measurement dates were 24.3, 10 and 88 days. All data gaps were linearly interpolated. Subsequently, regular quasi weekly time series were generated by selecting readings of the $1^{st}$, $8^{th}$, $15^{th}$ and $22^{th}$ of each month of the 20 years period. This resulted in a set of 141 groundwater head time series each with readings at the same 960 quasi weekly measurement dates. Thus, the last "quasi-week" exhibited different lengths for the different months.

The observation wells were irregularly distributed (Figure 1). This is a consequence of the mission of the state office to monitor possible anthropogenic influences on groundwater level, thus focusing on more densely populated areas. At 35 sites wells were screened at different depths. Distances between closest observation wells ranged from 0 km, at the sites with

several wells, to 20.2 km, with a mean distance between closest wells of 3.4 km. At each site, the screens were numbered from the surface downwards (Figure 1). Different numbers of the screens do not necessarily imply different aquifers. Comparing two observation wells of different sites, a higher screen number does not imply that the distance of the screen or of the water level to the surface is larger as well (Figure 2). The mean depth to groundwater, measured as distance of
180  groundwater head to the cap of the well head during the observation period, ranged from 0.65 m to 30.96 m, with a mean of 6.30 m and a standard deviation of 5.36 m. The distribution of monitored mean depths to groundwater was heavily skewed towards smaller depths (Figure 2). Due to the complex hydrogeological setting, the capture zones of the wells usually are not known with sufficient detail.

As a rough approximation of the autocorrelation of the time series in spite of the not completely regular weekly sampling
185  intervals the correlation of the series with itself shifted by one time step were assessed, yielding a correlation coefficient of r = 0.97 with a standard deviation of 0.04. The correlation in the spatial domain between the series of closest adjacent observation wells was substantially weaker with a mean correlation of r = 0.76 and exhibited substantially larger variability with a standard deviation of 0.24.

190  **Figure 2 Mean depths to groundwater at the observation wells in m of the complete set of series and separately for each screen number.**

## 3 Methods

### 3.1 Principal component analysis

195  The spatiotemporal variability in the groundwater head monitoring network was summarized with linearly independent principal components (PCs) determined by principal component analysis (PCA). The PCs are derived by an eigenvalue decomposition of the covariance matrix of the analysed set of variables, that is, the observed groundwater head series. To achieve equal weighting of all series, we applied PCA to the z-scaled groundwater head series, thus each series was scaled to zero mean and standard deviation of one. This corresponds to performing PCA with the correlation matrix of the
200  groundwater head series.

Each PC consists of an eigenvalue, eigenvector (loadings) and scores. The size of the eigenvalue of a PC in relation to the sum of all eigenvalues of all PCs corresponds to the share of variance in the data set that is assigned to that PC. The loadings

are the weights in the linear combination of the analysed variables, here the z-scaled groundwater head series, to calculate the scores of the PCs.

205 In this study, the analysed data set consisted of a set of time series all covering the same period and exhibiting the same dates. Thus, the scores of the PCs are times series with the same dates as the analysed time series. Please note, that this condition is a requirement for the suggested application of the reference hydrograph in this study. The Pearson correlation coefficient was used to describe the relationships between the observed groundwater head time series and that of a selected PC. This yielded for each PC a characteristic spatial pattern of "occurrence" of the respective PC time series at the

210 observation wells. The Pearson correlation values of this relationship correspond to the spatial pattern of the loadings of a PC multiplied with the square root of the eigenvalue of the respective PC. For better readability of the results we used here the Pearson correlation values for the maps of the loadings. Thus, each PC is associated with a characteristic temporal pattern (time series of the scores) and spatial pattern (loadings).

We performed PCA with the function "prcomp" of the default package "stats" in R version 3.4.1 (R Core Team, 2017). For
215 more details on PCA, please see Jolliffe (2002).

### 3.2 Stability of PCs

Being a data driven approach, the PCA results depend on the selection of data. Thus, for the assessment of the typical regional behaviour it is crucial to use only those PCs which are rather insensitive to the specific selection of analysed
220 observation wells, i.e. complete series, and to the specific selection of measurement dates. Those PCswere considered to be representative for the monitored region in the studied period.

To assess the stability of spatial patterns of the PC loadings on the scale of the network, we performed 10,000 PCAs based on random subsamples of the 960 quasi-weekly measurement dates and compared the PC loadings of the different runs. We calculated the squared Pearson correlation coefficient ($R^2$) of all combinations of loadings of PC 1 of the different PCA runs,
225 all combinations of loadings of PC 2 of the different PCA runs, etc. for all PCs with eigenvalue larger than one. This so called Kaiser criterion is a common threshold to select in case of PCA of z-scaled variables only those PCs which summarize more variance of the data set than one of the analysed variables (Jolliffe, 2002). The whole stability analysis was performed with subsamples of 70% of all measurement dates. Analogously, we performed the stability analysis of temporal patterns of the PC scores with 10,000 PCAs each based on random selections of 70% of the 141 complete groundwater head series. We
230 considered only those PCs as stable of which the correlations of both, the spatial as well as the temporal patterns, exhibited a median $R^2 > 0.9$.

### 3.3 Well-specific reference hydrograph and residuals

The focus of this study was to present an approach to quickly screen all groundwater head series in a comprehensive
235 monitoring network for problems with data quality and anthropogenic effects. For this purpose each observed hydrograph wa decomposed into a "normal" part describing the typical behaviour for the monitored region and the well- specific deviations from it. The "normal" behaviour at each observation well was estimated by a well-specific reference hydrograph. The well-specific reference hydrograph was calculated by multiple linear regression of the observed series with the time series of the scores of the stable PCs. The residuals from the regression (residuals), which is the part of the series that has not been
240 assigned to the reference hydrograph, describe the local deviations from the normal behaviour at each observation well. In contrast to many other approaches, it is not required that the residuals are white noise or alike. Systematic structures in the residuals like drifts, trends, cyclic patterns, sudden shifts or distinct periods of deviations indicate that the respective pattern

is not representative for the whole data set, but is a local peculiarity instead. This serves as indication for technical problems or anthropogenic influence.

245    The squared linear correlation coefficient $R^2$ of the observed versus the respective reference hydrographs was used for a first assessment which observation wells exhibit rather normal behaviour and which do not. This enabled to rank all wells in the network quickly according to their "normality". Those wells at which the observed series were not well represented by the respective reference hydrograph, i.e. those which exhibited low $R^2$ with the reference hydrograph and high $R^2$ with the residuals, were considered the most promising candidates to exhibit anthropogenic effects or measurement errors. The

250    residuals were visually examined for peculiarities. Those were than discussed with the experts from the LUNG in the context of knowledge on the local conditions and other information available for the respective observation wells to derive hypotheses on the causing drivers.

## 4 Results

255    ### 4.1 Stability of PCs

The first ten PCs of the PCA of the complete data set exhibited eigenvalues larger than one. Thus in the following we present the results of the first ten PCs only. The comparison of the spatial patterns of the PCAs which were performed with the sub datasets based on the random subsampling of the measurement dates is summarized in Figure 3a. In general, the median correlation between the loadings of the PCs of the different PCAs was decreasing and the variability of correlation between

260    the loadings was increasing with increasing rank of the PCs (Figure 3b). All PCs except the 8th and the 10th exhibited notably stable spatial patterns with a median correlation of $R^2 > 0.9$.

The comparison of the temporal patterns of the PCAs which were performed with the sub datasets based on the random selections of observation wells, i.e. complete series, is summarized in Figure 3b. Generally, with increasing rank of the PCs the median correlation between the scores of the PCs of the different PCAs was decreasing and the variability of correlation

265    between the scores was increasing (Figure 3b). Only the first four PCs exhibited notably stable temporal patterns with a median correlation of $R^2 > 0.9$.

Accordingly, the first four PCs were considered stable on the scale of the network, accounting for 80.8% of the observed variance in the groundwater head series. The variance assigned was for PC 1: 48.3%, PC 2: 17.2%, PC 3: 9.5% and PC 4: 5.8%. The assigned temporal and spatial patterns of the complete data set with all 141 series are shown in Figure 4 and

270    Figure 5. In the following the analysis is restricted to these four stable PCs.

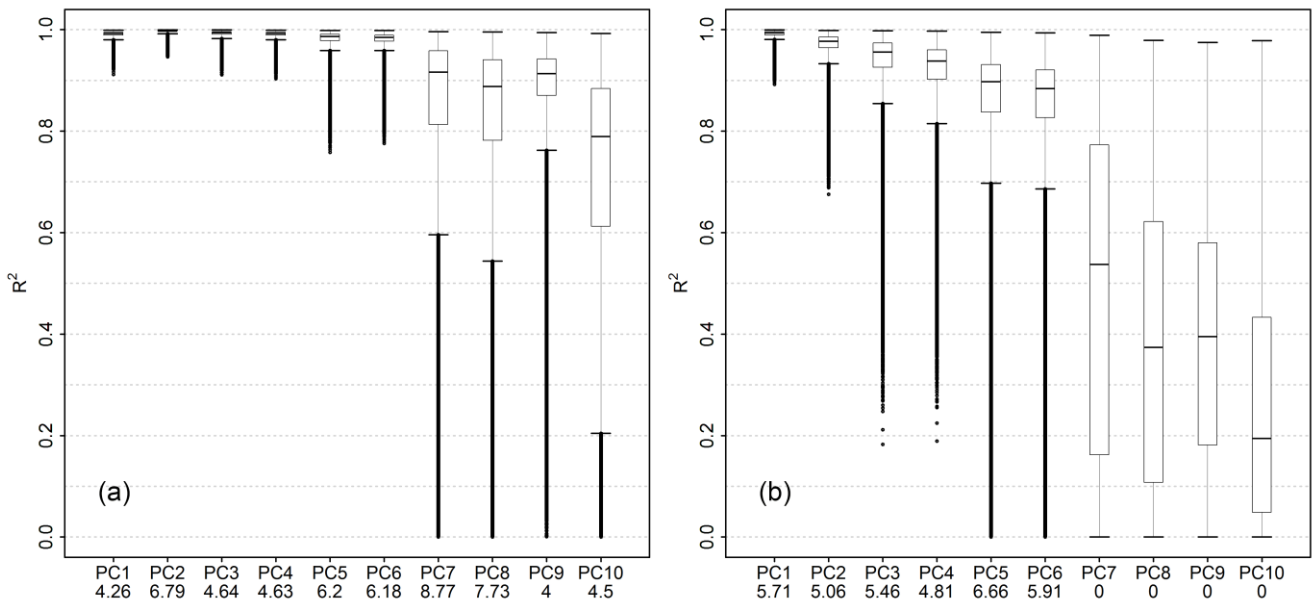



**Figure 3: Comparison of the PCA results of the stability analysis. (a) Correlation of loadings based on the random subsampling of the measurement dates to assess the stability of spatial patterns. (b) Correlation of scores based on the random selection of observation wells, i.e. complete series, to assess the stability of temporal patterns. The boxes indicate the quartiles, the whiskers all dates which are within the range of the first quartile - 1.5\* the interquartile range, respectively the third quartile + 1.5\* the interquartile range. Percentage of values outside the whiskers of a boxplot is given in the labelling of the x-axis for each PC.**

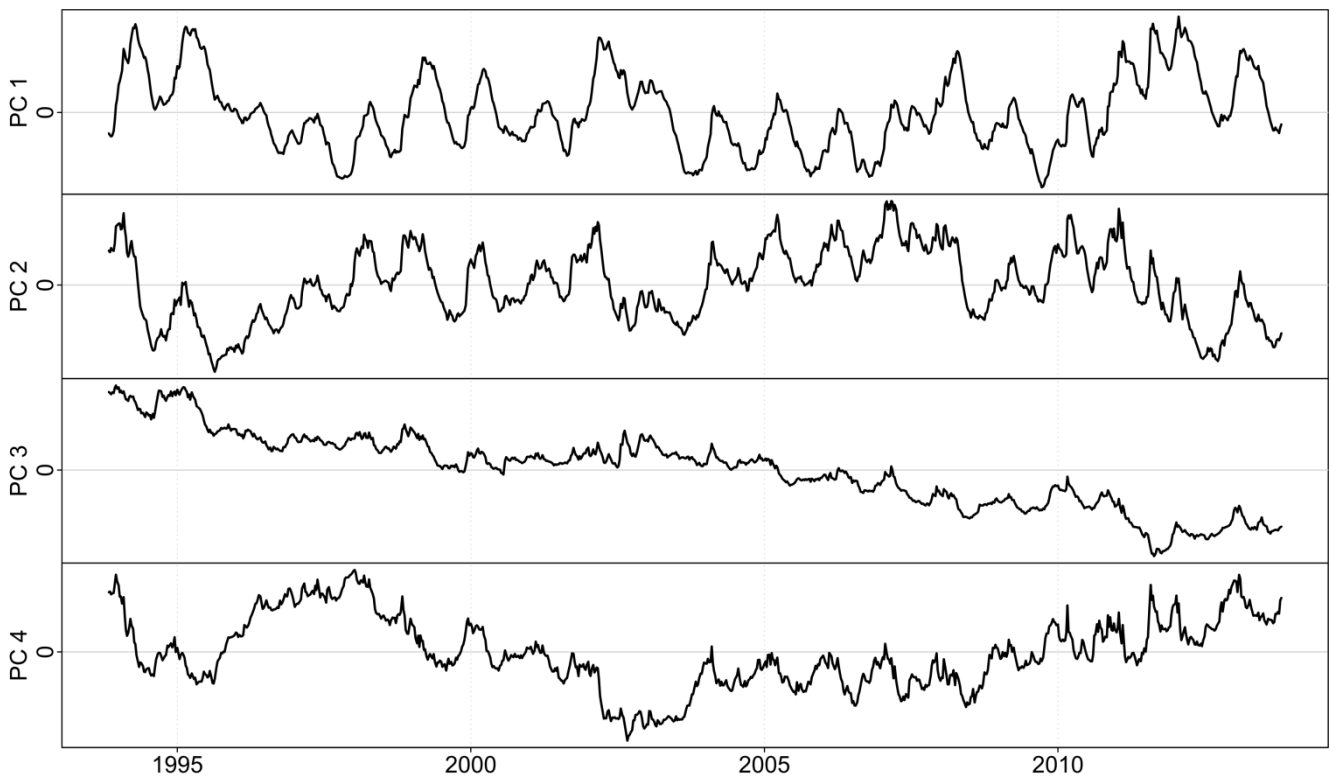



**Figure 4: Time series of scores of the stable PCs 1 to 4.**

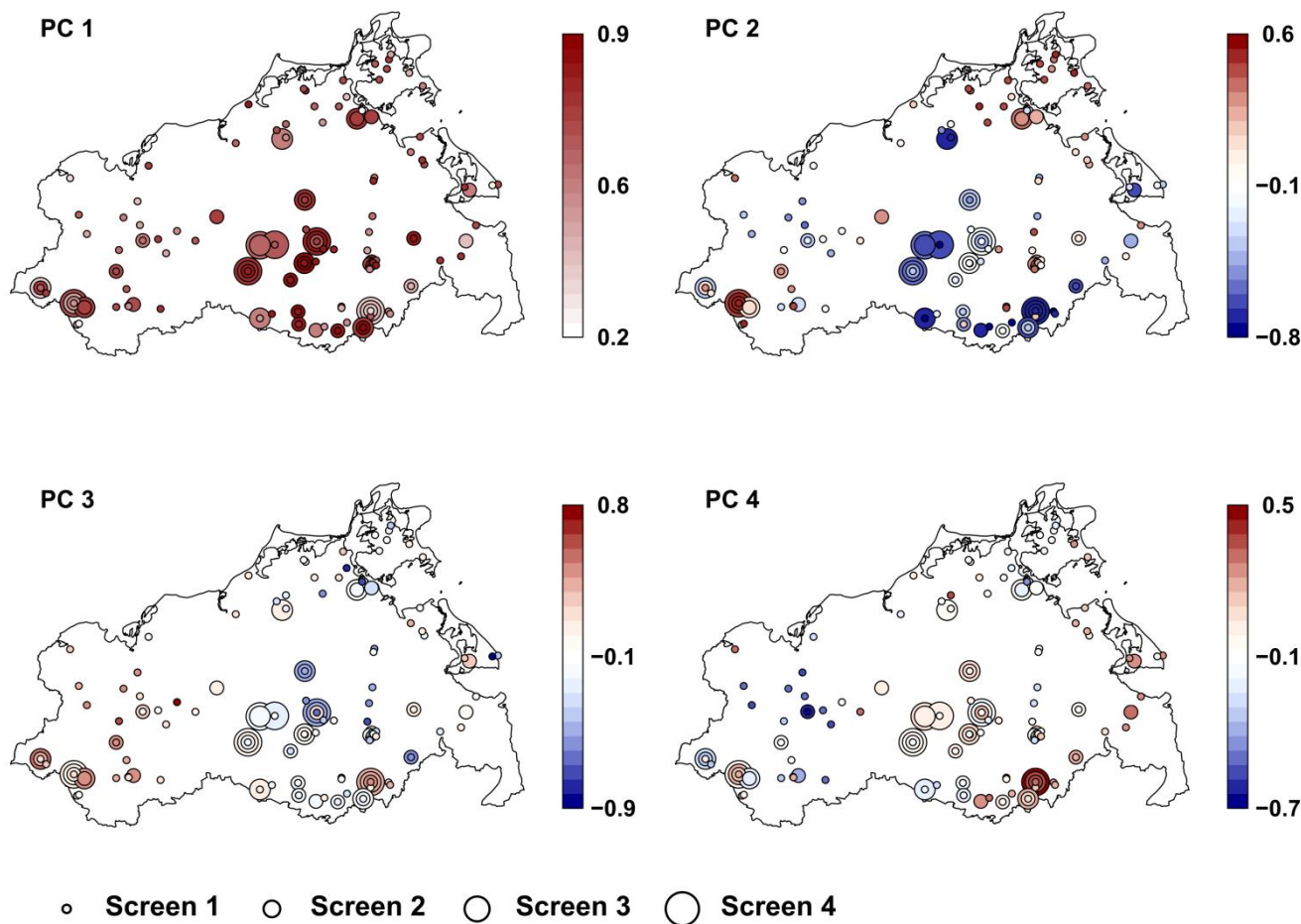



Figure 5: Spatial patterns of loadings of the stable PCs 1 to 4.

## 4.2 Well-specific reference hydrograph and residuals

285   At each well, the residuals of the reference hydrograph were checked for peculiarities. We present here two examples. At well Deven (Figure 6) the observed hydrograph was in general very well represented by the well-specific reference hydrograph ($R^2 = 0.84$). The mean depth to groundwater was 9.52 m. A single period stuck out in the plot of the residuals. At the end of October 1998 the residuals showed a sudden shift of approximately 17 cm towards higher water level, followed by a similar sudden shift "back to the old level" in December 1999. Those shifts were not clearly identifiable in the observed

290   time series itself nor in the reference hydrograph.

Observation well Neubrandenburg UP exhibited in general a relatively good fit of observed series and reference hydrograph ($R^2 = 0.77$) with the exception of two anomalous periods in 1997-1998 and 2007-2008, a series of minor deviations before 1997 and another relatively strong deviation in 2011 (Figure 7a and c). The mean depth to groundwater was 4.61 m. For comparison we considered the close-by observation well NB-Hotel Vier Tore approximately 600 m further south which was

295   formerly excluded from the PCA due to known anthropogenic influence (section 2.2). The mean distance of well head to groundwater level was 3.74 m. Here we calculated the reference hydrograph in the same manner as multiple linear regression of the observed hydrograph with PCs 1–4. Again the observed series exhibited in general a relatively good fit with the reference hydrograph (Figure 7b and c). However, compared to observation well Neubrandenburg UP the two periods of strong deviation in 1997-1998 and 2007-2008 were more pronounced in the residuals and, in contrast to observation well

300   Neubrandenburg UP, clearly visible in the observed series as well. This was reflected in a substantially weaker correlation between the observed series and the reference hydrograph ($R^2 = 0.46$).

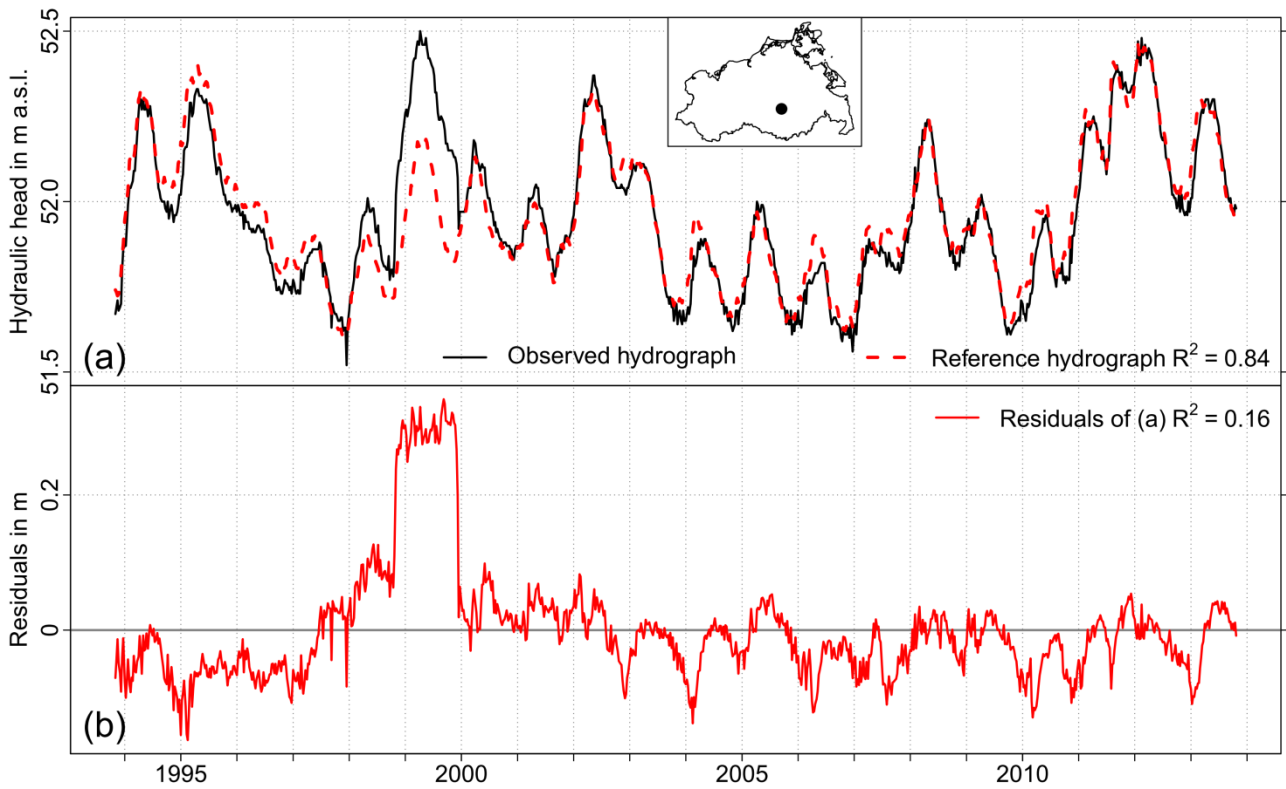



Figure 6: (a) Time series of hydraulic head and the reference hydrograph of well Deven. (b) Time series of residuals. Location of
305   the well is marked as black dot in the inset map. Height of the well head at well Deven is 61.48 m a.s.l.. Correlation of the observed
series with the reference hydrograph and with the residuals is given as $R^2$.

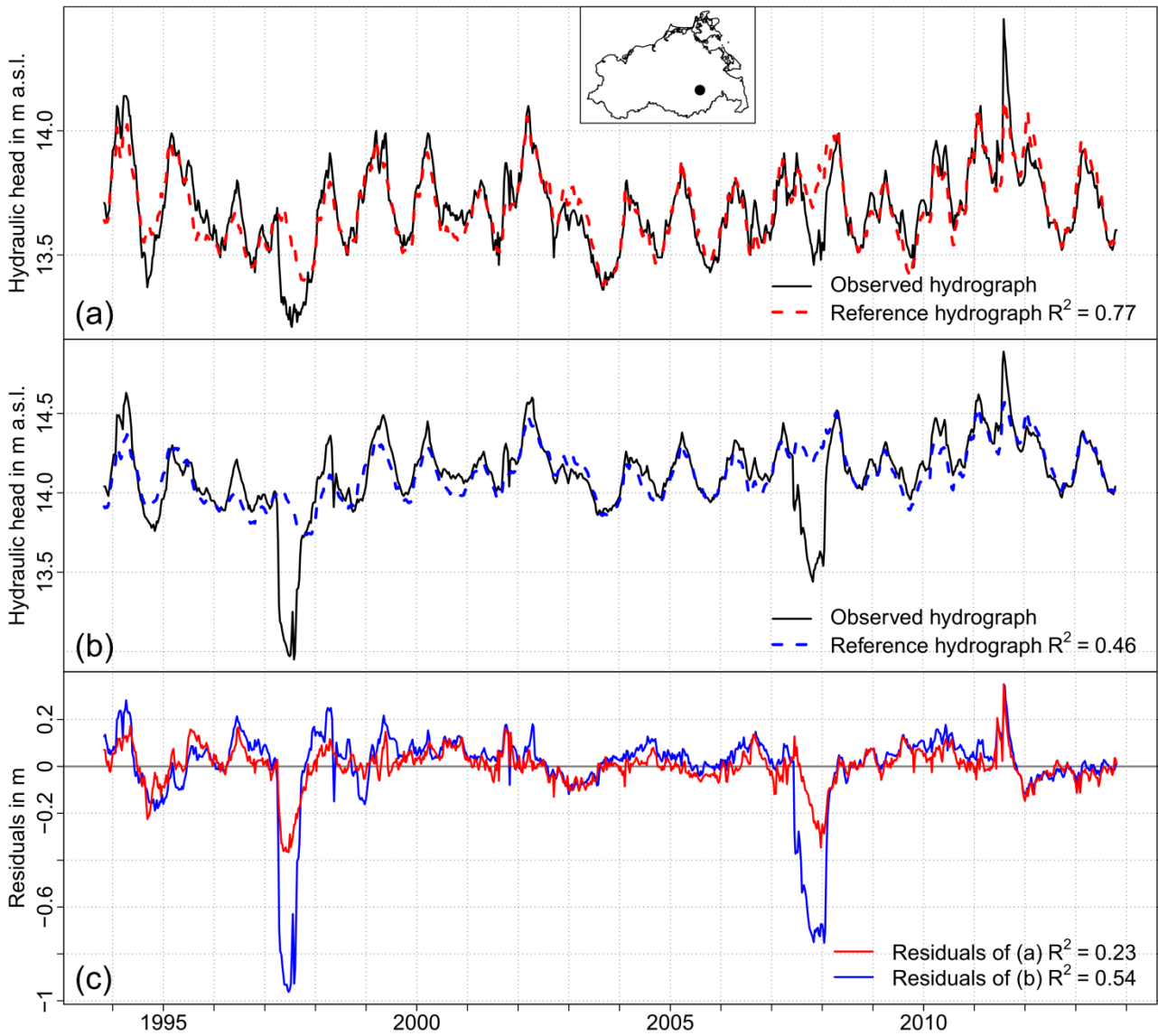



**Figure 7 Time series of hydraulic head and the reference hydrograph of (a) well Neubrandenburg UP and (b) well NB-Hotel Vier Tore. (c) Time series of residuals. Location of the wells is marked as black dot in the inset map. Height of the well head at well Neubrandenburg UP is 18.3 m a.s.l., at well NB-Hotel Vier Tore 17.86 m a.s.l.. Correlation of the observed series with the reference hydrograph and with the residuals is given as $R^2$.**

## 5 Discussion

### 5.1 Stability of PCs

To select only those PCs which are representative for the monitored region in the analysed period a series of PCAs were performed based on randomly selected subsets of the complete data set to identify the stable PCs. The stable PCs are those of which the assigned spatial patterns (loadings) and temporal patterns (scores) were rather insensitive to the selection of analysed observation wells and measurement dates. Only the stable PCs were considered for the further analysis.

Earlier studies which used PCA to summarize hydrological variability in a region analysed the stability of their results in a similar way (Smirnov, 1973; Lins 1985a). Those attempts were limited to the comparison of a few PCAs, respectively a few different configurations of the data set. The correlation analysis in this study extended the assessment of stability of PCs towards random selections of the analysed data in both the spatial and the temporal domain, and a substantially larger number of different configurations of the analysed data set.

The comparison of the results of the 10,000 variants of PCA revealed that the stability of individual PCs decreased in general with decreasing rank of the PCs (Figure 3). Clear differences between the stability of temporal and spatial patterns of the PCs were observed. PCA results were more sensitive to changes in the selection of considered observation wells (Figure 3b) than to changes in the selection of measurement dates (Figure 3a). This likely reflects the stronger mean correlation between subsequent observations at the sites (autocorrelation) compared to the mean correlation of the series of closest adjacent sites (spatial domain) (section 2.2). The first four PCs were found to be stable (Figure 3). This gave some confidence that their spatial and temporal patterns were indeed characteristic for the monitored region in the analysed period and not restricted to the specific selection of observation wells or measurement dates. Compared to the established Kaiser criterion (section 3.2) the number of considered PCs was reduced from ten to four.

## 5.2 Well-specific reference hydrographs and residuals

The well-specific peculiarities in the time series of the residuals of the reference hydrographs might be of different origin. First of all they might be caused merely by technical problems. For example we interpreted the single period in 1999 which sticks out in the plot of the residuals of observation well Deven as a step-wise shift of the logger (Figure 6b). This shift was in phase with the seasonal pattern of the observed series and was therefore not obvious from the visual inspection of the observed series alone, while it was clearly visible in the residuals.

An example for well-specific peculiarities in the residuals due to local anthropogenic influence is given in Figure 7. Observation well NB-Hotel Vier Tore was excluded from the PCA because its hydrograph was known to be influenced by the lowering of groundwater level due to construction works of an underground car park in approximately 100 m distance in 1997-1998 and those of another underground car park in approximately 200 m distance in 2007-2008. While this influence was clearly visible in the observed series at that observation well (Figure 7b), it was not obvious in the observed series at observation well Neubrandenburg UP, especially not the second deviation in 2007-2008 (Figure 7a). This is most probably because Neubrandenburg UP was further away from both construction sites, namely approximately 400 m each. However, in the residuals both periods became clearly visible for both observation wells although to different degrees at the two wells (Figure 7c).

Such clear assignment of anthropogenic influence to a local deviation from the regional behaviour is only possible if the scale of the respective effect is rather local in comparison to the scale of the monitoring network as a whole, and the scale of spatiotemporal resolution of the network in particular. The latter enables a distinct localization of the influence. An anthropogenic effect which induces similar groundwater head dynamics at a substantial amount of the series of the data set would be incorporated in the leadings PCs (Wriedt et al., 2017), and thus would affect the reference hydrographs. However, such an effect is hardly likely. Moreover, the presented approach does not differentiate anthropogenic from "natural" effects per se. Rather, it decomposes the time series into regional patterns which can be assigned to many or all time series of the data set and local patterns which are restricted to a few or single sites. Another restriction is that PCA based on z-scaled series considers only temporal patterns of groundwater head, but ignores the absolute values. Thus, it does not allow any inferences whether the observed groundwater head on average is higher or lower than it would be under natural conditions. Then again, the z-scaling of the series makes the analysis robust against systematic differences or discontinuities in the spatial distribution of mean groundwater recharge.

The spatial clustering of observation wells indicated a spatial bias of the monitoring (Figure 1). It reflected the focus of the monitoring on anthropogenic water use, for example close to settlements, towns, etc., which is prerequisite for sustainable water management. Because all the series were equally weighted by z-scaling (section 3.1), the derived PCs and consequently the determined normal behaviour was distorted towards areas with higher density of observation wells (Karl et al., 1982) as well as towards smaller mean depths to groundwater (Figure 2). This should be considered for any interpretation

of the reference hydrographs as well as of their local deviations. However, in this study, the stable PCs used for the calculation of the reference hydrographs turned out to be rather robust with respect to the selection of observation wells (section 4.1), suggesting that the results were not primarily determined by the local clustering of the observation wells.

In general, the reference hydrograph is a relatively good-natured and robust PCA application for mainly two reasons. First, the selection of the considered PCs is transparent and reproducible. The approach prevents the consideration of PCs which exhibit pronounced instability of the associated spatial and temporal patterns, for example PCs with "degenerate eigenvalues", that is eigenvalues which are indistinguishable within their range of uncertainties (Hannachi et al., 2007). Second, not the single PCs are used, but the combination of the stable PCs. Thus, it is not necessary to interpret single PCs as drivers of groundwater head variability or distinct processes. Consequently, describing the regional behaviour with the reference hydrograph is also applicable in cases in which the interpretation of single PCs is severely hampered, for example if the associated spatial patterns of the PCs mainly reflect the shape of the analysed domain (domain shape dependence) (Buell, 1975, 1979; Richman, 1986, 1993).

Some PCA applications involve rotation of the considered PCs to achieve more simple structures, i.e. more localized spatial patterns, which might support the interpretation of single PCs (Lins, 1985b; Richman, 1986; Jolliffe, 1987; Jolliffe, 2002; Hannachi et al., 2007). It is possible to combine such applications with the reference hydrograph application. For the suggested screening application rotation of the PCs does not change the results as long as the rotation is performed with all stable PCs or only with a subset of the stable PCs. In this case the reference hydrographs and their residuals are the same whether they are calculated from the rotated or the un-rotated PCs. Concerning the reference hydrographs the decisive question is which PCs are included in the calculation.

The presented approach uses the spatiotemporal variability in a large set of groundwater head series to determine individual reference hydrographs for each observation well. Thus, it does neither require the identification of clusters of similar wells or single reference wells, nor assumptions on the catchments of the wells, hydraulic connection between the wells, etc. This is in contrast to approaches which select some of the monitored wells as reference observation wells being representative for the whole monitoring network or for subgroups or sub-regions only. For example other PCA applications used the clustering of observation wells in the scatterplot of loadings of PC 1 versus PC 2 to identify "index" wells for each cluster (Winter al., 2000) or applied PCA directly to subgroups of a monitoring network, which were determined based on an estimation of the physical relatedness of the observation wells before, to identify firstly the "principal wells" of the subgroups and subsequently the most representative wells for the whole network by ranking all wells according to their number of occurrences as principal well (Gangopadhyay et al., 2001).

However, despite the different approaches of how to determine the most representative observation wells it has to be considered that observation wells which are of little representative value for the whole monitoring network might be of high informative value for example with respect to anthropogenic influence specific to single observation wells. In contrast to the selection of single wells which are considered especially representative or atypical for the network, the suggested approach in this study enables to order all wells in a network quickly according to their representativeness and yields for each well an estimation of the local well-specific behaviour.

## 5.3 Other applications

In addition to the presented applications in this study, other applications of the reference hydrographs and residuals are possible. One option is to use the series of the stable PCs (scores) as predictor variables in a linear regression to fill gaps in hydraulic head series which were not analysed with the PCA but which exhibit some overlap with the monitoring period covered by the PCA. For example the reference hydrograph of observation well NB-Hotel Vier Tore (Figure 7b) was calculated although the groundwater head series was not part of the PCA. In this study it was merely used for identification

of the influence of the construction works, but it could be used to replace the periods which were influenced by the construction works with an estimation of unaffected groundwater head dynamics as well (Figure 7b). In the spatial domain, calculating the Pearson correlation of excluded series with the stable PCs for limited overlapping time periods can be used to extend the spatial coverage of the loadings maps. For both applications, the random subsampling analysis can be used to estimate how many missing values in the series might be tolerable. For this data set we would be confident to consider series with up to 30% missing values during the twenty-year period. The maximum time gap in this study was 88 days (section 2.2) and most of the time gaps were substantially smaller (Figure S2). Thus, the results are most likely less stable for a few long gaps compared to numerous but short gaps.

Another application is to identify distinct reference observation wells by selecting those observation wells at which the correlation between the reference hydrograph and the observed series is above a certain threshold, for example $R^2 > 0.9$. Those would be considered as most representative observation wells for the whole monitoring network. In case the period covered by the series of the reference observation wells exceeds the period covered by the PCA, they can be used as predictor variables in a linear regression to extrapolate the series of the stable PCs scores.

The methods to extend the spatial and temporal coverage of the PCs should be handled with care. However, because only the stable PCs were used, there should be no major bias, as long as the extension is performed only for some years or small numbers of additional series. If new data are available, that cover a larger area or a longer period, it is in general preferable to perform a new PCA with all available data to account for systematic changes in the temporal dynamics of the analysed groundwater system, and systematic changes in the monitoring network geometry and spatial distribution of the observation wells.

The detection of changes in characteristic temporal patterns (scores) and their occurrence in the monitoring network (maps of loadings) between different observation periods is another application of the reference hydrographs and their residuals. It has to be noted that a direct comparison of the temporal patterns of two periods can only be performed for temporally overlapping periods. For non-overlapping periods, a comparison is restricted to general time series characteristics, for example the timing or amplitude of a seasonal pattern.

## 6 Conclusions

The application of the reference hydrographs and their residuals proved to be a straightforward way to quality check the data and to identify candidates for local anthropogenic influence. Both applications are actually an interpretation of the temporal dynamics of local anomalies in the observed groundwater head series. In contrast to other approaches the identification of those local anomalies is based on the correlation among the observed groundwater head series only and not based on physical models or empirical relationships with any predictors of groundwater head dynamics. The assignment of local anomalies to the residuals is not restricted to specific types of temporal patterns. The residuals merely comprise what cannot be ascribed to the reference hydrographs by means of the stable PCs. This can be short term structures like sudden shifts or distinct periods of deviation as well as long term structures like drifts, trends or cyclic patterns. The presented approach also does not require an interpretation of single PCs as distinct physical processes or functional relationships. This limitation with respect to direct physical interpretation of the results brings with it the benefit that the only information required are series of (groundwater) hydraulic head readings all measured at the same instants of time. Other suggested applications for the stable PCs are for example data-driven gap-filling in the observed series, spatial and temporal extrapolation of the reference hydrographs or the identification of distinct reference observation wells. The approach to determine the stable PCs is transparent and reproducible.

The computational demand is very low. Time series of the well-specific deviations from normal behaviour (residuals) are easily derived and enable a fast screening for well-specific peculiarities. In monitoring practice, ranking rank all wells in the

network according to their "normality" can be used to distribute resources, in particular to support the decision which observation wells, i.e. which series, should be investigated in more detail. Especially for larger data sets it can be helpful to apply time series tools like filters to identify for example step changes or trends in the analysis of the residuals. Those tools can be easily incorporated in the workflow once the residuals are derived. However, based on our experience, we advocate that visual inspection should be included in the quality check workflow as well. Furthermore the residuals can be used to categorize the deviations from the normal behaviour. Thus, we recommend the presented approach as a fast screening tool for the assessment of comprehensive groundwater monitoring networks.

**Author contribution**

CL performed the formal analysis and visualization, carried out the research, and prepared the manuscript and the replies to the reviewers with contributions from GL. GL acquired the funding, designed the initial research concept and supervised the project. CL and GL developed the methodology in the course of the project.

**Competing interests**

The authors declare that they have no conflict of interest.

**Data availability**

Up to date groundwater head series of the observation wells used in this study are provided on request by the State Office of Environment, Nature Conservation and Geology (LUNG), Güstrow, Germany.

**Acknowledgements**

The first author received funding by the State Office of Environment, Nature Conservation and Geology (LUNG) of the federal state of Mecklenburg-Vorpommern. We thank Beate Schwerdtfeger and Heike Handke from LUNG for the initialisation of the project, the provision of the data and the productive collaboration throughout the project, Christoph Merz, Philipp Rauneker and Katharina Brüser from Leibniz Centre for Agricultural Landscape Research (ZALF), Marcus Fahle from the Federal Institute for Geosciences and Natural Resources (BGR) and Tobias Hohenbrink from the Technical University of Berlin (TU Berlin) for the fruitful discussions, Berry Boessenkool for the inspiration for Fig. S1, Klaus-Dieter Fichte, Toralf Henke and Michael Höft from the State Office of Agriculture and Environment Vorpommern (STALUP), as well as Hanjo Polzin and Peter Schuldt from the Mining Office of Mecklenburg-Vorpommern in Stralsund, Konrad Höppner from CEMEX Germany, Torsten Abraham from Hydro-Geo-Consult GmbH and Jacob Möhring from LUNG for expert knowledge on different observation wells, Heiko Hennig from Umweltplan Greifswald and Toralf Hilgert from HGNord for information on former projects with similar data sets in the study region, and Andreas Mitschard from LUNG for the continuous technical support.

# References

Bartlein, P. J.: Streamflow anomaly patterns in the U.S.A. and southern Canada — 1951–1970, Journal of Hydrology, 57, 49-63, doi: 10.1016/0022-1694(82)90102-0, 1982.

Buell, C. E.: The topography of the empirical orthogonal functions, Preprints, Fourth Conference on Probability and Statistics in Atmospheric Sciences, 188-193, 1975.

Buell, C. E.: On the physical interpretation of empirical orthogonal functions, Preprints, Sixth Conference on Probability and Statistics in Atmospheric Sciences, 112-117, 1979.

Coppola, E., Szidarovszky, F., Poulton, M. and Charles, E.: Artificial Neural Network Approach for Predicting Transient Water Levels in a Multilayered Groundwater System under Variable State, Pumping, and Climate Conditions, Journal of Hydrologic Engineering, 8, 348-360, doi: 10.1061/(ASCE)1084-0699(2003)8:6(348), 2003.

Coulibaly, P., Anctil, F., Aravena, R. and Bobde, B.: Artificial neural network modeling of water table depth fluctuations, Water Resources Research, 37, 885-896, doi: 10.1029/2000WR900368, 2001.

EU, Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy, Official Journal of the European Communities, 1-70, 2000.

GADM, Database of Global Administrative Areas - Maps, http://www.gadm.org, version 2.8, last access: 2017-11-06, 2017.

Gangopadhyay, S., Gupta, A. D. and Nachabe, M.: Evaluation of Ground Water Monitoring Network by Principal Component Analysis, Groundwater, 39, 181-191, doi: 10.1111/j.1745-6584.2001.tb02299.x, 2001.

Gleeson, T., Alley, W. M., Allen, D. M., Sophocleous, M. A., Zhou, Y., Taniguchi, M. and VanderSteen, J.: Towards Sustainable Groundwater Use: Setting Long-Term Goals, Backcasting, and Managing Adaptively, Groundwater, 50, 19-26, doi: 10.1111/j.1745-6584.2011.00825.x, 2011.

Gottschalk, L.: Hydrological regionalization of Sweden, Hydrological Sciences Journal, 30, 65-83, doi: 10.1080/02626668509490972, 1985.

Hannah, D. M., Smith, B. P. G., Gurnell, A. M. and McGregor, G. R.: An approach to hydrograph classification, Hydrological Processes, 14, 317-338, doi: 10.1002/(SICI)1099-1085(20000215)14:2<317::AID-HYP929>3.0.CO;2-T, 2000.

Hennig, H., Hilgert, T., Kolbe, R. and Lückstädt, M.: Optimierung des Grundwasserstandsmessnetzes Mecklenburg-Vorpommern - Eine Methodik auch für andere Flächenländer, Hydrologie und Wasserbewirtschaftung, 46, 117-124, 2002.

Hennig, H. and Hilgert, T.: Dränabflüsse – Der Schlüssel zur Wasserbilanzierung im nordostdeutschen Tiefland, Hydrologie und Wasserbewirtschaftung 51, 248-257, 2007.

Hilgert, T.: Grundwasserneubildung Mecklenburg-Vorpommern - Aktualisierung und ergänzende Beschreibung, internal project report, 2009.

Hilgert, T. and Hennig, H.: Grundwasserfließgeschehen Mecklenburg-Vorpommerns -- Geohydraulische Modellierung mit Detrended Kriging, Grundwasser, 22, 17-29, doi: 10.1007/s00767-016-0348-6, 2017.

Hirsch, R. M.: A Perspective on Nonstationarity and Water Management, JAWRA Journal of the American Water Resources Association, 47, 436-446, doi: 10.1111/j.1752-1688.2011.00539.x, 2011.

Hodgson Frank D., I.: The Use of Multiple Linear Regression in Simulating Ground-Water Level Responses, Groundwater, 16, 249-253, doi: 10.1111/j.1745-6584.1978.tb03232.x, 1978.

Hohenbrink, T. L., Lischeid, G., Schindler, U. and Hufnagel, J.: Disentangling the Effects of Land Management and Soil Heterogeneity on Soil Moisture Dynamics, Vadose Zone Journal, 15, doi: 10.2136/vzj2015.07.0107, 2016.

Hurst, H.: Long Term Storage Capacities of Reservoirs, Transactions of the American Society of Civil Engineers, 116, 776-808, 1951.

Jolliffe, I.: Principal Component Analysis, Second Edition, Springer, 2002.

Karl, T. R., Koscielny, A. J. and Diaz, H. F.: Potential Errors in the Application of Principal Component (Eigenvector) Analysis to Geophysical Data, Journal of Applied Meteorology, 21, 1183-1186, doi: 10.1175/1520-0450(1982)021<1183:PEITAO>2.0.CO;2, 1982.

Koutsoyiannis, D.: Nonstationarity versus scaling in hydrology, Journal of Hydrology, 324, 239–254, doi: 10.1016/j.jhydrol.2005.09.022, 2006.

Lewandowski, J., Lischeid, G. and Nützmann, G.: Drivers of water level fluctuations and hydrological exchange between groundwater and surface water at the lowland River Spree (Germany): field study and statistical analyses, Hydrological Processes, 23(15), 2117–2128, doi: 10.1002/hyp.7277, 2009.

Lins, H. F.: Interannual streamflow variability in the United States based on principal components, Water Resources Research, 21, 691-701, doi: 10.1029/WR021i005p00691, 1985a.

Lins, H. F.: Streamflow Variability in the United States: 1931–78, Journal of Climate and Applied Meteorology, 24, 463-471, doi: 10.1175/1520-0450(1985)024<0463:SVITUS>2.0.CO;2, 1985b.

Lins, H. F.: Regional streamflow regimes and hydroclimatology of the United States, Water Resources Research, 33, 1655-1667, doi: 10.1029/97WR00615, 1997.

Longuevergne, L., Florsch, N. and Elsass, P.: Extracting coherent regional information from local measurements with Karhunen-Loève transform: Case study of an alluvial aquifer (Rhine valley, France and Germany), Water Resour. Res., 43, WO4430, doi: 10.1029/2006WR005000, 2007.

LUNG, Map of the distance to the surface of the border between fresh and saline groundwater, german: "Tiefenlage Süß-/Salzwassergrenze", dl_versalzung.shp, available at: https://www.umweltkarten.mv-regierung.de/script/, last access: 2018-06-04, 1984.

LUNG, Map of groundwater recharge, german: "Grundwasserneubildung", gwn.shp, available at: https://www.umweltkarten.mv-regierung.de/script/, last access: 2017-06-27, 2009.

Mandelbrot, B. and Van Ness, J.: Fractional Brownian Motions, Fractional Noises and Applications, SIAM Review, 10, 422-437, doi: 10.1137/1010093, 1968.

Mandelbrot, B. B. and Wallis, J. R.: Some long-run properties of geophysical records, Water Resources Research, 5, 321-340, doi: 10.1029/WR005i002p00321, 1969.

Manhenke, V., Reutter, E., Hübschmann, M., Limberg, A., Lückstadt, M., Nommensen, B., Peters, A., Schlimm, W., Taugs, R. and Voigt, H.-J.: Hydrostratigrafische Gliederung des nord- und mitteldeutschen känozoischen Lockergesteinsgebietes, Zeitschrift für Angewandte Geologie, 47, 146-152, 2001.

Marchant, B., Mackay, J. and Bloomfield, J.: Quantifying uncertainty in predictions of groundwater levels using formal likelihood methods, Journal of Hydrology, 540, 699–711, doi: 10.1016/j.jhydrol.2016.06.014, 2016.

Martens, S. and Wichmann, K.: In: Lozán, J. L.; Graßl, H.; Hupfer, P.; Karbe, L. & Schönwiese, C.-D. (Ed.), Global Change: Enough water for all? – 2.11 Groundwater salinisation, 2011.

R Core Team: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/ (last access: 8 October 2017), 2017.

Rajaee, T., Ebrahimi, H. and Nourani, V.: A review of the artificial intelligence methods in groundwater level modeling, Journal of Hydrology, 572, 336–351, doi: 10.1016/j.jhydrol.2018.12.037, 2019.Richman, M. B.: Rotation of principal components, Journal of Climatology, 6, 293-335, doi: 10.1002/joc.3370060305, 1986.

Richman, M. B.: Comments on: 'The effect of domain shape on principal components analyses', International Journal of Climatology, 13, 203-218, doi: 10.1002/joc.3370130206, 1993.

Sahoo, S., Russo, T. A., Elliott, J. and Foster, I.: Machine learning algorithms for modeling groundwater level changes in agricultural regions of the U.S., Water Resources Research, 53, 3878-3895, doi: 10.1002/2016WR019933, 2017.

Skøien J., O., Blöschl, G. and Western A., W.: Characteristic space scales and timescales in hydrology, Water Resources Research, 39, doi: 10.1029/2002WR001736, 2003.

Smirnov, N.: Asynchronous long-period streamflow fluctuations in the European U.S.S.R, Soviet Hydrology: Selected Papers, 46-49, 1972.

570   Smirnov, N.P.: Spatial patterns of long-period streamflow fluctuations in the European U.S.S.R, Soviet Hydrology: Selected Papers, 112-122, 1973.

Statistikportal, Statistical Offices of the federal republic and the federal states of Germany, https://www.statistikportal.de/de/bevoelkerung/flaeche-und-bevoelkerung, last access: 2018-11-15, 2018.

Triki, I., Trabelsi, N., Hentati, I. and Zairi, M.: Groundwater levels time series sensitivity to pluviometry and air temperature:
575   a geostatistical approach to Sfax region, Tunisia, Environmental Monitoring and Assessment, 186, 1593-1608, doi: 10.1007/s10661-013-3477-8, 2014.

Upton K., A. and Jackson C., R.: Simulation of the spatio-temporal extent of groundwater flooding using statistical methods of hydrograph classification and lumped parameter models, Hydrological Processes, 25, 1949-1963, doi: 10.1002/hyp.7951, 2011.

580   Von Asmuth, J. R., Maas, K., Bakker, M. and Petersen, J.: Modeling Time Series of Ground Water Head Fluctuations Subjected to Multiple Stresses, Groundwater, 46, 30–40, doi: 10.1111/j.1745-6584.2007.00382.x, 2008.

Winter, T., Mallory, S., Allen, T. and Rosenberry, D.: The Use of Principal Component Analysis for Interpreting Ground Water Hydrographs, Groundwater, 38, 234-246, doi: 10.1111/j.1745-6584.2000.tb00335.x, 2000.

Wriedt, G.: Verfahren zur Analyse klimatischer und anthropogener Einflüsse auf die Grundwasserstandsentwicklung,
585   Grundwasser, 22, 41-53, doi: 10.1007/s00767-016-0349-5, 2017.

Wunsch, A., Liesch, T. and Broda, S.: Forecasting groundwater levels using nonlinear autoregressive networks with exogenous input (NARX), Journal of Hydrology, 567, 743–758, doi: 10.1016/j.jhydrol.2018.01.045, 2018.