Hydrological Processes

WILEY

RESEARCH ARTICLE OPEN ACCESS

Unravelling Regional Water Balance Dynamics in Anthropogenically Shaped Lowlands: A Data-Driven Approach

Jenny Kröcher[1,2] | Gohar Ghazaryan[1,3] | Gunnar Lischeid[1,2]

[1]Leibniz Centre for Agricultural Landscape Research (ZALF), Müncheberg, Germany | [2]Institute of Environmental Sciences and Geography, University of Potsdam, Potsdam, Germany | [3]Geography Department, Humboldt-University of Berlin, Berlin, Germany

Correspondence: Jenny Kröcher (jenny.kroecher@zalf.de)

Received: 30 September 2024 | **Revised:** 11 December 2024 | **Accepted:** 19 December 2024

Funding: This research is part of the BMBF Land-Innovation-Lausitz project 'Monitoring der Landschaftsraumoberflächen und Biodiversität mittels Drohnen- und Satellitendaten in der Modellregion Lausitz' and was supported by the German Federal Ministry of Education and Research (BMBF, Grant number 03WIR3020B).

Keywords: groundwater dynamic | lowland | principal component analysis | satellite-based evapotranspiration

ABSTRACT

Effective management of water resources in anthropogenically shaped lowlands requires a comprehensive understanding of hydrological processes and balancing various effects in complex settings, especially like lowland hydrology. Unlike mountainous headwater catchments with shallow soils, lowland hydrology is typically dominated by groundwater dynamics, often exhibiting pronounced spatial correlation lengths, though other factors may also contribute. This necessitates consideration of distant anthropogenic impacts in water resources management. This study focuses on the Lusatia region in the northern German, a lowland area heavily altered by mining activities, including extensive groundwater lowering and rebound, impacting the overall water regime. We applied an efficient, data-based approach to unravel various impacts on the landscape water balance over a 30-year period (1993–2022). We integrated over 1800 ground-based time series data on groundwater levels, surface water dynamics and runoff, supplemented by evapotranspiration estimates from multi-temporal Landsat satellite data to account for land use effects. Through principal component analysis, we identified key patterns driving water balance dynamics. The first four components explained 84% of the variance in groundwater and surface water levels, as well as of runoff dynamics. The dominant processes attributed to these components include anthropogenic influences from mining activities, as well as natural hydrogeological effects such as seasonal variability and the damping of the groundwater recharge signal in the unsaturated zone. A separate principal component analysis that included evapotranspiration data explained 87% of the variance, with the first component predominantly reflecting seasonal variations and subsequent components elucidating land use impacts and long-term vegetation changes. By linking both analyses, we generated comprehensive maps detailing the spatial distribution of effects on regional water balance. Our approach provides a quantitative tool to assess the size and influence of natural and anthropogenic effects on water resources, offering a comprehensive tool for assessing spatial and temporal effects on hydrological dynamics in a lowland region affected by human activities.

1 | Introduction

Monitoring and understanding the water balance at the landscape scale is crucial for effective water resources management, especially in the context of climate change and increasing anthropogenic pressure. Traditionally, these efforts have relied on ground-based observations of groundwater and surface water (Condon et al. 2021; Kumar et al. 2021a). While these methods provide high temporal resolution of hydrological dynamics, they fall short in offering a comprehensive view of the water balance at the landscape scale. To extrapolate point data to area-wide information, modelling approaches are widely used (Barthel and Banzhaf 2016; Kumar et al. 2021a, 2021b). However, these models require extensive datasets, limiting their feasibility for widespread landscape monitoring (Barthel and Banzhaf 2016; Kumar et al. 2021b). Additionally, models often struggle to detect and analyse the underlying factors influencing water balance dynamics, particularly in complex landscapes with significant anthropogenic alterations or where landscape management practices have been implemented. This limitation can result in significant uncertainties (Kumar et al. 2021b).

Current methods for transferring groundwater time series or analysing water dynamics from well-monitored sites to areas without direct observations typically rely on spatial similarity approaches and regression analysis (Heudorfer et al. 2019; Rinderer, van Meerveld, and McGlynn 2019; Giese et al. 2020; Haaf et al. 2023). These methods assume that observed patterns in one location can be applied to similar conditions elsewhere (Barthel et al. 2021) but, so far, have not been applied to heterogenous landscapes with massive anthropogenic modifications (Haaf et al. 2023), such as the Lusatia region. This approach has substantial limitations particularly in complex real-world settings with various synchronous effects which differ in their respective spatial correlation length (Blöschl and Sivapalan 1995). For example, weather effects are fairly homogeneous at a scale of some 10^2 to 10^4 km in the lowlands, whereas open pit mining affects water levels at a 10^0 to 10^1 km scale, and local groundwater extraction for irrigation at much smaller scale (Blöschl and Sivapalan 1995; Hangen-Brodersen, Strempel, and Grünewald 2005; Longuevergne, Florsch, and Elsass 2007).

The Lusatia region, shaped by extensive mining activities, has experienced significant impacts on water resources and soil fertility over the past decades (Gerwin et al. 2023). With the region facing both climate change and socio-economic shifts, a comprehensive landscape transformation is imminent. Future challenges include the flooding of open-cast mines, the rebound of lowered groundwater levels, the development and preservation of fertile soils and the protection and revitalisation of peatlands (Hangen-Brodersen, Strempel, and Grünewald 2005; Krümmelbein et al. 2012). Additionally, the region is likely to experience further water management difficulties due to increasing air temperature and evapotranspiration (ET) and decreasing precipitation (Reyer et al. 2012), posing challenges across various scales of landscape transformation.

To effectively address these challenges, innovative and efficient landscape monitoring approaches are essential (Biemelt, Schapp, and Grünewald 2011). Advances in remote sensing and statistical analysis offer new opportunities to monitor water dynamics on high spatial and temporal scales. We propose a principal component analysis (PCA) that assists in mapping of spatial and temporal effects on water dynamics, incorporating measured time series of water levels and area-wide ET data obtained from remote sensing. PCA is a statistical method used to identify underlying spatial effects by analysing large sets of time series variables (Bretherton, Smith, and Wallace 1992). This method is increasingly applied in hydrology to determine the factors influencing hydrological and hydrogeological time series (Gottschalk 1985; Page et al. 2012; Yue et al. 2020; Meggiorin et al. 2022). In northeastern Germany, PCA has been widely used to analyse time series of groundwater head and surface water runoff at the federal states (NUTS2) level (Thomas et al. 2012; Lehr and Lischeid 2020; Lischeid et al. 2021).

Despite its proven value, PCA has not been applied in regions with extensive anthropogenic interventions, such as Lusatia. This research aims at elucidating the spatial and temporal patterns of various effects and at generating a spatio-temporal, area-wide estimation of individual effects to be related to anthropogenic or natural processes, respectively. Importantly, this estimation will be based solely on existing and available measurement data, including ground-based and remote sensing datasets, without requiring any additional data directly related to the sources of these effects (e.g., time and amount of groundwater lowering in mining areas). This approach is proposed as a very efficient way contributing to a comprehensive understanding of the hydrological dynamics in complex landscapes.

In particular, the underlying research questions are:

1. What are the key spatio-temporal patterns in ground-based time series on groundwater level and discharge, and how accurately can we differentiate here between anthropogenic and natural effects?
2. What are the key spatio-temporal patterns in ET time series, and how do these patterns correlate with specific land-use types and climatic factors in the Lusatia region?
3. To what extent can the integration of groundwater, surface water and ET time series data effectively map the primary factors influencing the water balance in both spatial and temporal dimensions?

2 | Data and Methods

2.1 | Study Site

The study region is located in northeastern Germany, approximately 80 km south of Berlin, and spans around 9820 km^2. It is situated in the Niederlausitz region, part of the northeastern German lowlands (Figure 1). The boundaries of the area are defined by the extent of the groundwater bodies within the region. The landscape is characterised by the Pleistocene glaciation with the typical glacial series of glacial valleys, outwash plains, terminal and ground moraines (Gerwin et al. 2023),

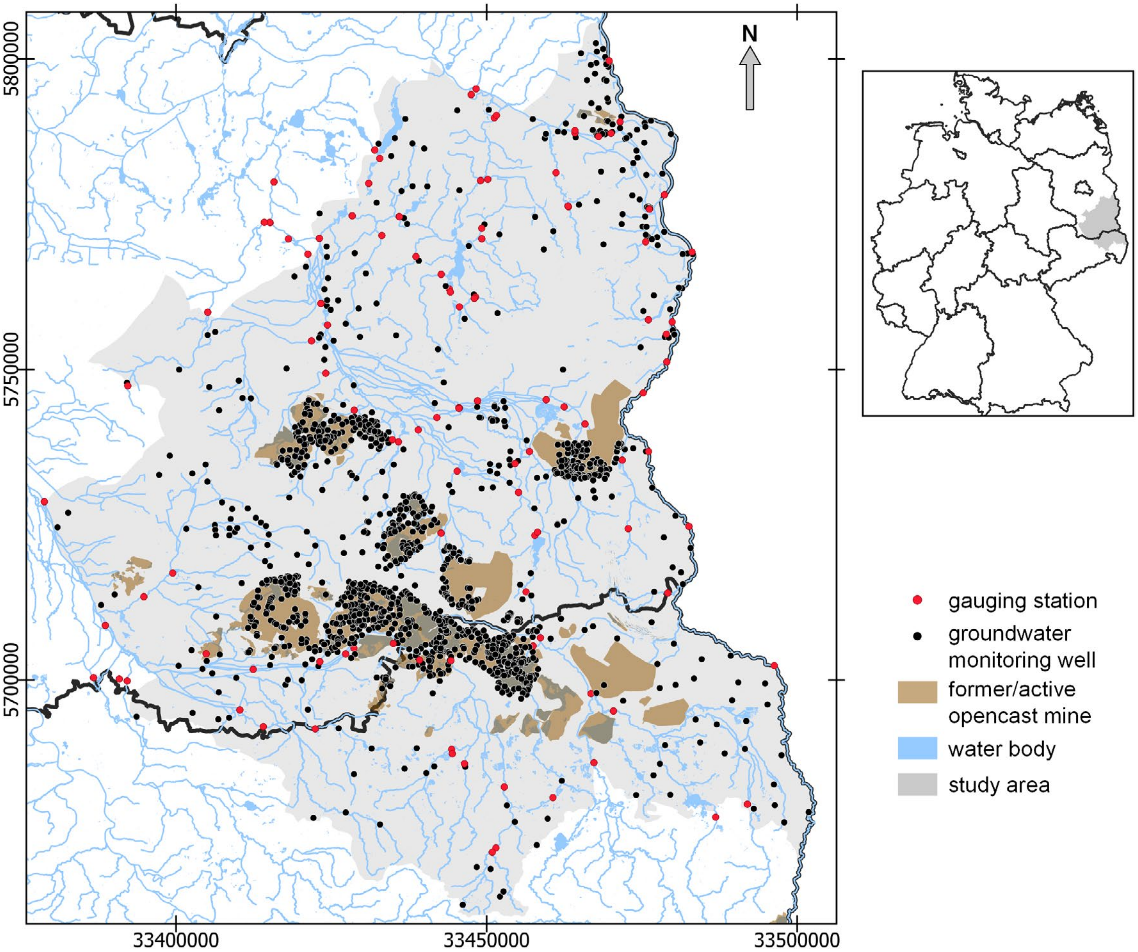


FIGURE 1 | Map of the study site in East Germany with former and active opencast mines and the selection of groundwater observation wells and gauging stations. Upper right map: Federal states of Germany with the location of the study region.

constituting a complex interplay of various porous aquifers. Only in the southernmost part of the study region fractured aquifers in Mesozoic and Palaeozoic bedrock are present.

The prevailing substrate of glacial sands, gravel, partly interspersed with clayey substrate, has a thickness of 40 to 120 m. Large groundwater resources have formed in the quaternary sediments. In the glacial valleys, shallow groundwater levels are characteristic and contributed to the formation of large fens in the Holocene.

Major rivers in the region include the Spree and Schwarze Elster (part of the Elbe catchment) and the Neisse (part of the Oder catchment). The stream network density varies from 0.3 km/km^2 on the plateaus of ground and terminal moraines to over 2 km/km^2 in the lowlands, particularly in the Spreewald, a large inland delta of the river Spree.

The landscape was shaped by opencast lignite mining below the Quaternary sediments (Gerwin et al. 2023) (Figure 1). The systematic extraction of lignite for over 120 years has created a footprint of over 906 km^2 (Statistik der Kohlewirtschaft 2024) and led to quantitative and qualitative impacts on water balance over an area of 3200 km^2 in Lusatia (Uhlmann et al. 2020). Groundwater extraction to access lignite seams, up to 120 m deep, has removed over 58 billion m^3 of water (Uhlmann et al. 2023). Open-cast mining led to the dumping of the sediment on the surface and large areas to be recultivated, on which agriculture and forestry are once again practised. On the other hand, some former open-cast mines were flooded with surface water from the largest streams in the region to prevent acidification caused by rising groundwater contaminated by pyrite oxidation (Hangen-Brodersen, Strempel, and Grünewald 2005).

The region's climate, with continental influences (Gädeke et al. 2017), has an annual average temperature of 9.7°C and 615 mm of precipitation based on a 30-year period from 1991 to 2020 (Deutscher Wetterdienst [DWD] 2023a, 2023b). The region's climatic water balance is negative at around −145 mm/year. on a long-term average from 1991 to 2020 (DWD 2024). Around half of the precipitation is distributed over the growing season from March to September, with mostly short, convective showers. Due to low water holding capacity of the prevailing sandy soils periods of low precipitation can lead to periods of drought (Krümmelbein et al. 2012; Reyer et al. 2012).

Land use is characterised by forestry and arable land (44% and 28% of the total area). Due to the sandy, low-yield soils and the occurring water stress, forestry use with pine (*Pinus sylvestris*) dominates as the predominantly planted tree species (Hofmann and Pommer 2005). Grassland is occasionally used in the glacial valleys on former moorland.

2.2 | Data

To assess hydrological patterns, we considered time series data from different water balance components. The analysis of groundwater, surface water levels and runoff from watercourses, which represents the water balance components runoff and groundwater and surface water storage, included time series datasets collected by conventional, ground-based monitoring. In total, data was provided from 16 177 measuring points, including 913 measuring points for surface water levels and discharges collected up to 2022.

We analysed ET as a further water balance component. For this purpose, we used the Landsat Collection 2 Provisional Actual ET Science Product (Senay 2018; Senay et al. 2023) due to its high spatial resolution of 30 m and temporal availability in the entire study period from 1993 to 2022. The data product is based on Landsat Collection 2 Level-2 Surface Temperature product which is the input to the Operational Simplified Surface Energy Balance (SSEBOp) model (Senay et al. 2022). Besides land surface temperature (LST), other inputs for the ET calculation include normalised difference vegetation index (NDVI) as well as air temperature, a digital elevation model (DEM), net radiation, and reference ET (Petrakis et al. 2024). Overall, we accessed and analysed 2459 ET images covering the entire study site.

We used additional environmental data such as land use, soil characteristics, geological and climatological data to interpret the results and to provide a comprehensive estimation of the effects on hydrological dynamics across the entire area. Table 1 lists the datasets and their sources. Datasets on land use were used to assess the impacts of land use distribution and transformation over nearly three decades, especially in relation to anthropogenic activities like mining, and their effects on the water balance. We derived lithological and hydraulic permeability data of the upper aquifer to describe the geological and hydrological determinants of groundwater flow and storage area. The soil's ability to retain water might affect both ET and groundwater recharge. Therefore, we included spatial soil water content data. Maps of groundwater table depth and mean annual precipitation provided additional key information on the spatial variability of groundwater levels and precipitation.

TABLE 1 | Data on hydrological and geological conditions, soil properties, and land use for estimation of hydrological dynamics.

Parameter	Dataset type	Description and data sources
Groundwater level, surface water level and runoff	Time series	Datasets requested from Brandenburg State Office for the Environment (LfU), the Saxony State Office for the Environment, Agriculture and Geology (LfULG), the Federal Institute of Hydrology (BfG) and the Lausitzer and Mitteldeutsche Bergbau-Verwaltungsgesellschaft mbH (LMBV)
Evapotranspiration	Raster dataset	Landsat Collection 2 Provisional Actual Evapotranspiration Science Product (Senay 2018; Senay et al. 2023)
Land use	ESRI-Shapefile	Derived from CORINE Land Cover/Land use data 2018 with 100 m spatial resolution (European Environment Agency 2019a) and aggregated to 8 land use classes (residential, open-cast mining, arable land, grassland, forest, moors and heathland, non-vegetated, water bodies)
Land use change	ESRI-Shapefile	Change in land use categories between 1990 and 2018 derived from CORINE Land Cover/Land Use data datasets (European Environment Agency 2019b, 2019c, 2019d, 2019e) and aggregated to 16 land use change classes based on the aggregated land use classes
Hydraulic permeability of the upper aquifer	ESRI-Shapefile	Derived from the General Hydrogeological Map of the Federal Republic of Germany 1:250 000 (HÜK250) of the German Federal Institute for Geosciences and Natural Resources (BGR) and the State Geological Surveys (SGD)
Lithology	ESRI-Shapefile	Derived from HÜK250 of the BGR and SGD
Soil water content	ESRI-Shapefile	Derived from the Soil Geological Map of Brandenburg at a scale of 1:300 000 (BÜK300) and the Soil Map of Saxony at a scale of 1:50 000 (BK50)
Groundwater table depth	ESRI-Shapefile	Derived from maps of groundwater table depth of LfU and LfULG
Mean annual precipitation	Raster dataset	30-year average annual precipitation based on station data, averaged for each year in the period from 1991 to 2020 and interpolated to a 1 km × 1 km grid (DWD Climate Data Center (CDC) 2021)

2.3 | Processing of Ground-Based and ET Time Series

The data analyses described below were carried out using R software version 4.2.3 (R Core Team 2024) and the packages *sf* (Pebesma 2018; Pebesma and Bivand 2023) and *dplyr* (Wickham et al. 2023) for time series and spatial data processing.

For the analysis of the ground-based time series for the 30-year period from 1993 to 2022, we initially selected time series automatically based on five criteria:

— Measurements begin before 1993,

— Measurements end after 2022,

— Average measurement intervals of at least 1 month,

— Gaps less than 1 year,

— Autocorrelation of the time series for the largest gap length is greater than 0.6.

The autocorrelation criterion serves as a measure of the temporal consistency and predictability of a time series. In combination with the maximum length of the gap, autocorrelation is an indicator of the extent to which the gaps in the time series can be reconstructed based on the previous data. We set the autocorrelation threshold of 0.6 based on visual inspection of various time series. We focused on the consistency of seasonal patterns, the stability of long-term trends, and the strength of the relationship between successive measurements. The aim of this inspection was to identify a threshold that retains predictable and meaningful temporal structures while excluding highly irregular time series. The measurement interval criterion was added to maintain seasonality in the subsequent interpolation of the gaps. After the criteria selection, a total of 590 time series remained, including 62 time series of surface water levels and discharges.

We synchronised the selected time series to weekly values in preparation for the PCA and filled existing data gaps using linear interpolation.

The processed time series were first z-normalised, ensuring that each series had a mean of zero and a standard deviation of one, which standardises the data and allows for a meaningful comparison across different series. Following this normalisation, we subjected the time series to PCA.

For the ET data, we processed the Landsat satellite images into time series per pixel for the entire period from 1993 to 2022. We checked the time series for completeness and removed time series with gaps of more than 1 year. Time series with smaller gaps were filled by linear interpolation followed by a synchronisation of all time series to weekly values. We included around 10 285 310 time series in the further analysis of ET time series.

2.4 | PCA

PCA of time series (also known as Empirical Orthogonal Function analysis) decomposes the time series into orthogonal (independent) principal components (PCs). These components are derived in such a way that each successive component captures the maximum possible variance from the data, given the constraints of being orthogonal to the preceding components. The eigenvalue of each PC indicates the amount of variance it explains. The sum of all eigenvalues equals the total variance in the data set. Therefore, the proportion of the explained variance by each component is determined by dividing its eigenvalue by the total sum of all eigenvalues.

The PCs are ranked by the amount of variance they explain, with the first component explaining the largest portion of the variance. We retained only those PCs that each explained more than 5% of the total variance, reducing the dimensionality of the dataset while preserving most of the important information.

For each PC, a separate time series was generated, where the length of each time series matches that of the input data. These time series represent the scores for each time step and each PC. The relationship between a PC and a time series of the input data was determined by calculating loadings. A loading indicates the strength and direction of correlation between each input time series and a specific PC. These loadings are crucial as they provide insight into how much each time series contributes to the identified patterns of variability (the PCs). High loadings on a particular component suggest that the corresponding time series shares a significant underlying effect or pattern, which may be related to specific physical processes influencing the time series at those measurement points. The loadings were calculated as the Pearson correlation between the z-normalised input time series and the time series of a specific PC.

In addition to the PCA of water level and discharge time series, we conducted a second PCA on Landsat ET time series to identify the most significant variables influencing this component of water balance.

2.5 | Mapping the Primary Factors Influencing the Water Balance

In the next step, we transferred the dominant effects on the groundwater and surface water balance, identified through the PCA of spatially point-based time series, to the entire study region. This allowed us to spatially delineate the effect strengths and make statements about the hydrological dynamics in the area. To achieve this, we predicted the loadings of the PCs from 10 area-wide available covariates. These covariates included the loadings of the PCA of ET time series and geological, land-use and soil characteristics (Table 1). Given the spatio-temporal and non-linear relationships between the effects on groundwater and surface water resources and area characteristics, we applied a random forest regression model to spatially predict these effects across the study area.

To increase training and test data samples, we included additional time series loadings by reconstructing incomplete time series previously excluded from the PCA. We selected 1253 time series (including 129 time series of surface water levels and discharge rates) for reconstruction, focusing on those with at least 25 years of data during the period from 1993 to 2022, and a minimum measurement frequency of four observations per year. Using the first eight PCs as predictors, we performed multiple linear regression to supplement the additional time series and calculate the loadings for each time series. We evaluated the quality of the regression models using the coefficient of determination R^2.

The loadings from the groundwater and surface water time series at the 1843 monitoring sites served as training (90%) and test data (10%) for model development and testing. Separate models were created for the area-wide regression of each PC's loadings. The random forest models were developed using the R software and the *caret* package (Kuhn 2008). We set the number of decision trees to 500, balancing estimation error optimization and computation time. The number of variables selected for each node in the decision tree was determined by the internal algorithm of the R function train(). This algorithm compares the model's accuracy across three different variable selections and chooses the number of variables that results in the highest model accuracy. We performed a 10-fold cross-validation to evaluate robust model performance metrics.

Additionally, the models were optimised following the approach of Millard and Richardson (2015). We used the variable importance calculated by the function train() and Spearman's rank correlation coefficient to determine pairwise correlations among the variables. Our goal was to use only the most relevant and uncorrelated variables for classification without increasing the error rate. The random forest algorithm was applied to the training data with 100 repetitions, creating a ranking of the five most relevant variables for each of the first to fourth PC. For the final models, we included only those variables that were within the five most relevant variables in at least one of the repetitions.

To evaluate the performance of the random forest regression model, we employed several accuracy metrics on the validation and testing dataset. These included the mean absolute error (MAE), root mean squared error (RMSE), and the R-squared (R^2) value, which collectively provide a comprehensive assessment of prediction accuracy and model fit.

3 | Results

3.1 | PCA of Ground-Based Time Series Data

PCA of the 590 selected ground-based time series produced 590 uncorrelated components. The first four components explained at least 5% of the variance each, specifically, 55%, 18%, 6% and 5%, respectively. Together, these four components explain approximately 84% of the total variance in the groundwater and surface water time series in Lower Lusatia. Consequently, the following analysis is restricted to these four PCs (PC_g).

The time series of scores of the first PC (PC_g1) shows an almost linear increase over the entire observation period with slight seasonal fluctuations (Figure 2). High positive loadings of PC_g1 are concentrated in areas corresponding to closed opencast mines, while high negative loadings are associated with active opencast mining sites (Figure 3). This suggests that PC_g1 captures the long-term effects of mining activities on groundwater trends, particularly the lowering and subsequent rebound of groundwater levels.

The second PC (PC_g2) exhibits pronounced seasonality, with clear fluctuations corresponding to seasonal hydrological changes (Figure 2). High loadings of PC_g2 occur at greater distances from the opencast mining areas (Figure 3), indicating that this component reflects natural hydrological seasonality, relatively unaffected by direct mining activities.

The time series of PC_g3, like PC_g1, displays low seasonality but with a marked increase between 1995 and 2011 (Figure 2). This period corresponds to the gradual recharge of the groundwater level after the termination of extensive drainage measures in former mining areas and illustrates the long-term hydrological recovery and its influence on the dynamics of the regional

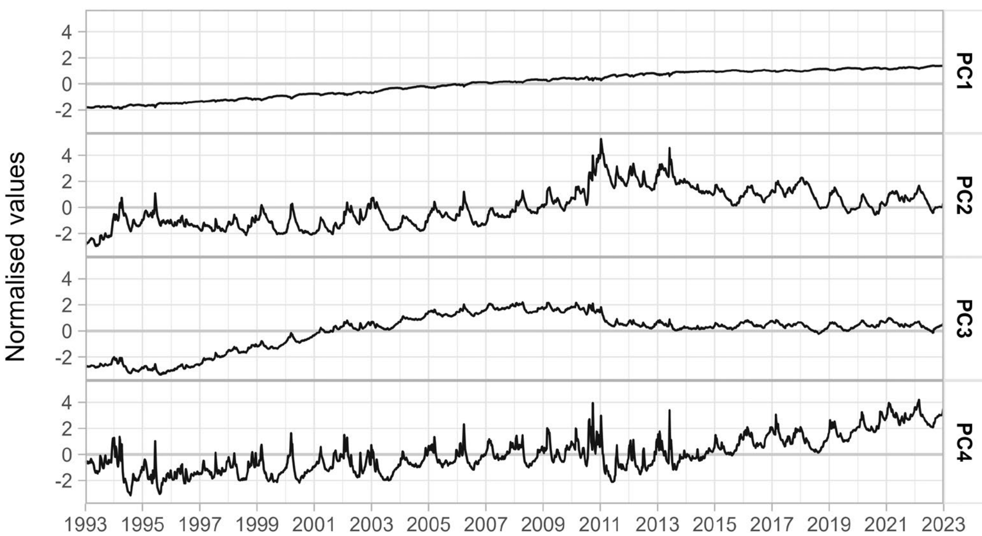


FIGURE 2 | Time series of scores of the first four principal components of the ground-based time series.

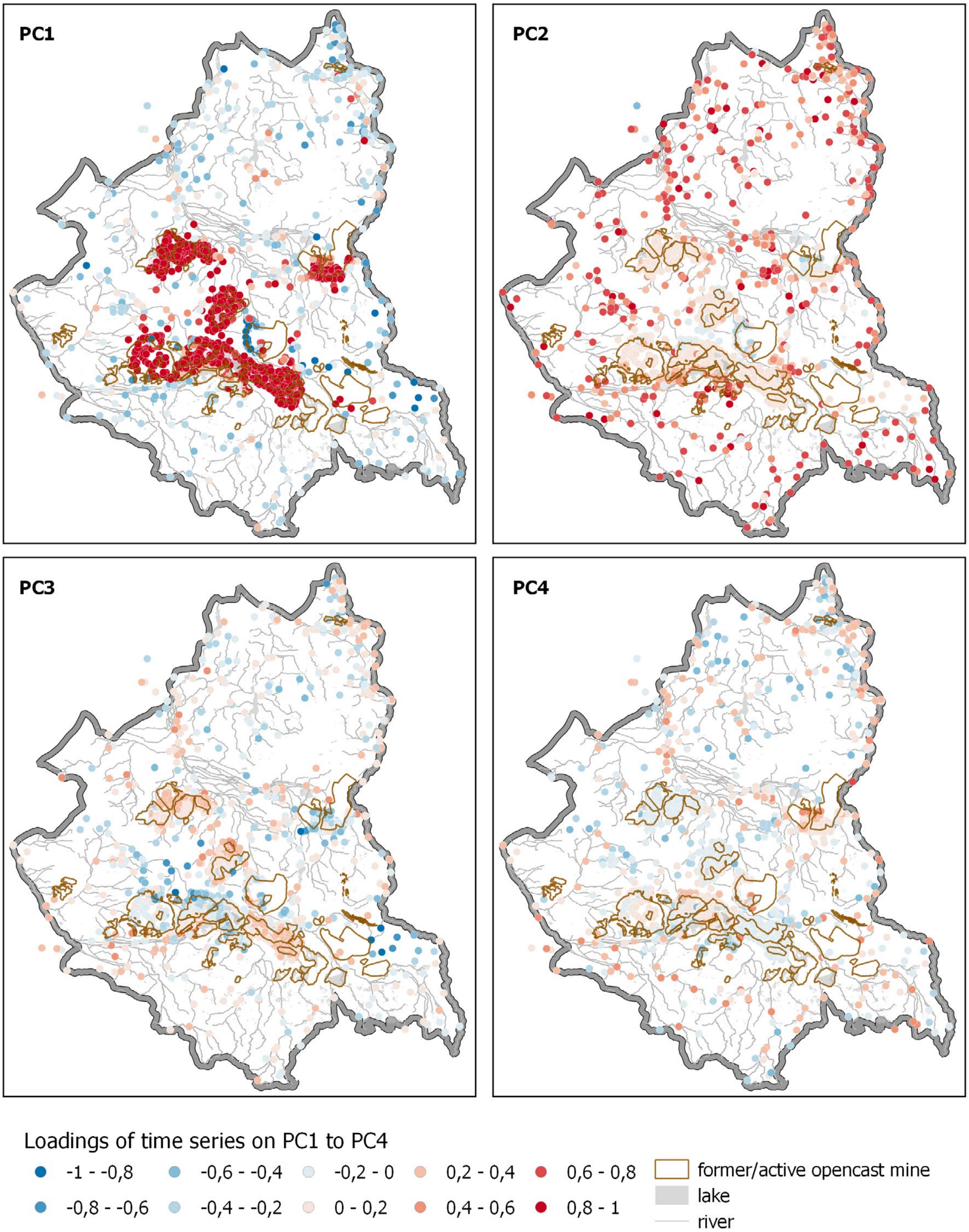


FIGURE 3 | Spatial patterns of loadings of PC_g1 to PC_g4 of ground-based time series.

water balance. The spatial distribution of PC_g3 loadings highlights areas with mining-influenced groundwater bodies that experienced both lowering and rebound over the past 20 years. High positive loadings specifically occur where, after the rebound, groundwater levels have been kept artificially lower than their natural state through pumping to prevent waterlogging and protect settlement infrastructure (Figure 3). PC_g3 thus captures delayed groundwater rebound effects linked to decommissioned mining operations.

When comparing the loadings of the PCs with various covariates, no significant correlations were found between the loadings of the first three PCs and land cover, geology, or soil type.

The fourth component (PC_g4) also shows seasonal patterns, but exhibits phase shifts compared to PC_g2, with variations in the timing and amplitude of peaks (Figure 2). The spatial patterns of PC_g4 are less distinct, but high positive loadings are generally found near watercourses. PC_g4 exhibited an inverse correlation

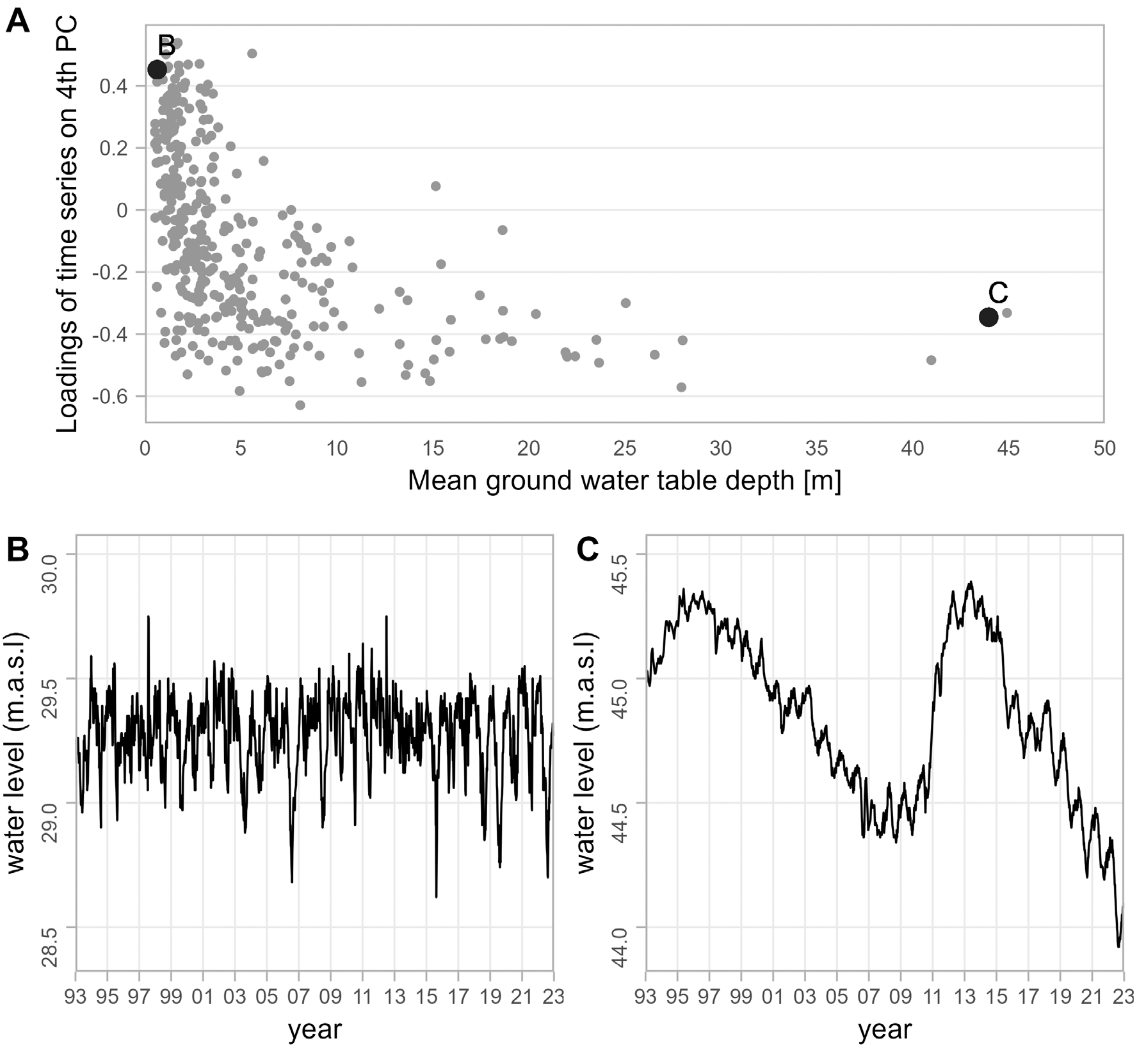


FIGURE 4 | Mean groundwater table depth and loadings of ground-based time series on PC_g4 (A) and time series of two monitoring sites with high (B) and low loadings (C).

with mean groundwater table depth, with loadings decreasing as groundwater table depth increased (Figure 4), suggesting this component represents attenuation of the groundwater recharge signal.

3.2 | PCA of ET Remote Sensing-Based Time Series Data

The first three PCs of PCA on ET data explain approximately 87% of the total variance in the time series data. Each of these components (PC_e) exhibits distinct seasonal patterns. However, no significant spatial correlations were found between their loadings and the 30-year mean precipitation pattern or soil characteristics.

PC_e1 accounts for 84% of the explained variance and captures the overall seasonal pattern of ET, characterised by high values in the summer months and low values in the winter months (Figure 5). The spatial distribution of PC_e1 loadings is relatively uniform across the study region. However, sparsely vegetated areas and opencast mines have lower loadings and therefore differ from the other areas (Figure 6). PC_e1 thus represents the dominant seasonal ET cycle across the study region.

PC_e2, explaining around 2% of the variance, reflects ET differences associated with land use. High positive loadings are seen in forestry sites and water bodies, while negative loadings are observed in agricultural areas and open-cast mines (Figure 6). The time series of PC_e2 shows an annual decrease in spring with negative peaks in May and June followed by a sharp increase and positive peaks in July and August (Figure 5). This component captures the impact of land use on ET patterns, particularly the earlier peak in agricultural areas compared to forests, as well as its sensitivity to dry years (e.g., 2018–2020).

The time series of PC_e3 shows a general increasing trend of ET peaks in summer, with less pronounced peaks during dry years (Figure 5). The spatial distribution of PC_e3 loadings reveals distinct patterns of high positive loadings in areas undergoing natural succession, afforestation and land use restoration (e.g., former military training sites and closed open-cast mines), and high negative loadings in active mining areas (Figure 6). PC_e3

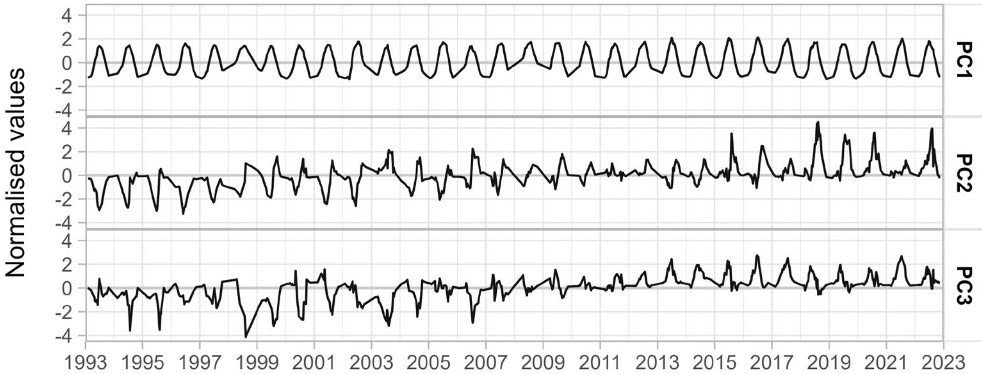


FIGURE 5 | Time series of scores of the first three principal components of the ET time series.

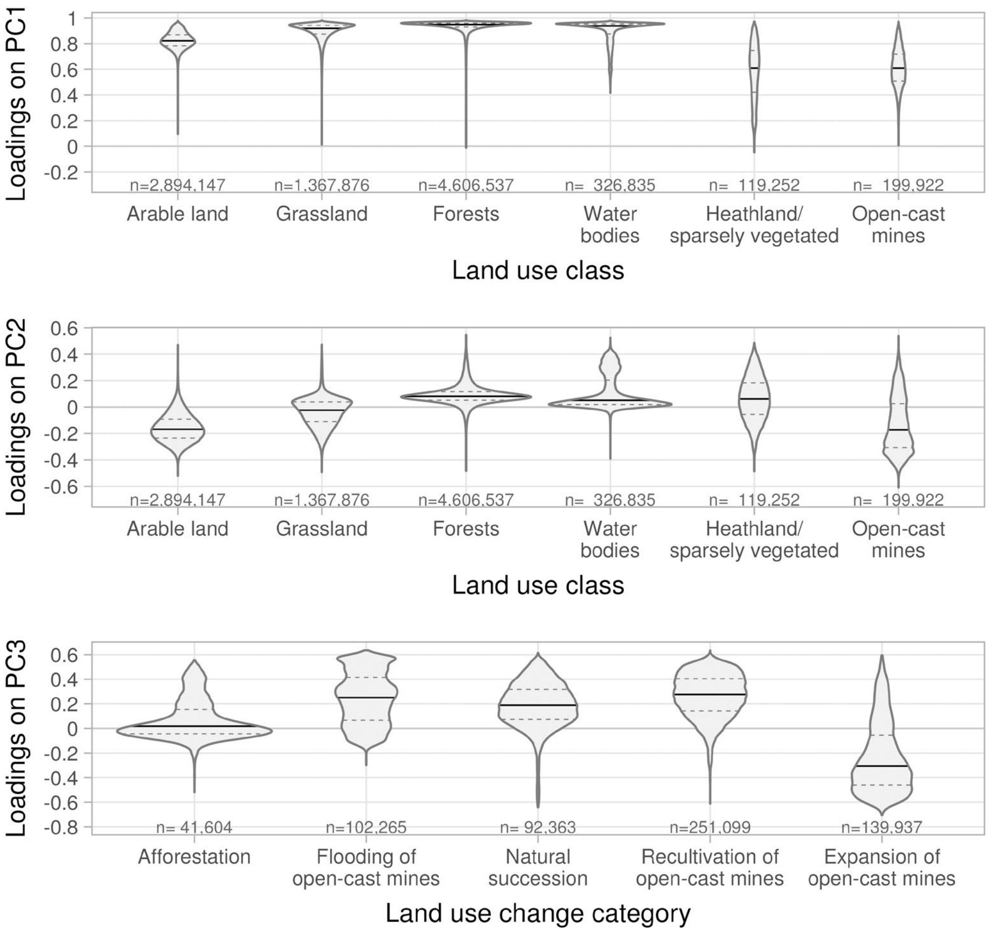


FIGURE 6 | Distribution of loadings on PC_e1 and PC_e2 divided into the most frequent land use classes and distribution of loadings on PC_e3 divided into the most frequent land use change categories in the period from 1993 to 2022.

therefore reflects long-term changes in ET due to land use transitions and restoration efforts over the past 30 years.

3.3 | Mapping of Primary Factors Influencing the Water Balance

To increase the training and test dataset for the random forest model, we added incomplete time series to the groundwater and runoff time series dataset. The additional time series were reconstructed using multiple linear regression. The quality of the 1253 reconstructed time series was evaluated using the coefficient of determination (R^2). For the reconstructed time series, R^2 values ranged from 0.2 to 0.99 (Figure 7), indicating varying levels of model performance. The median R^2 was 0.96, reflecting a high overall accuracy in the reconstruction process. This suggests that the majority of the reconstructed time series closely match the observed data and PCs, with the most linear regression models explaining a significant portion of the variance. However, the lower end of the R^2 range (0.2) indicates that some time series exhibited poorer fits, likely due to limitations in data coverage or underlying variability not captured by the model.

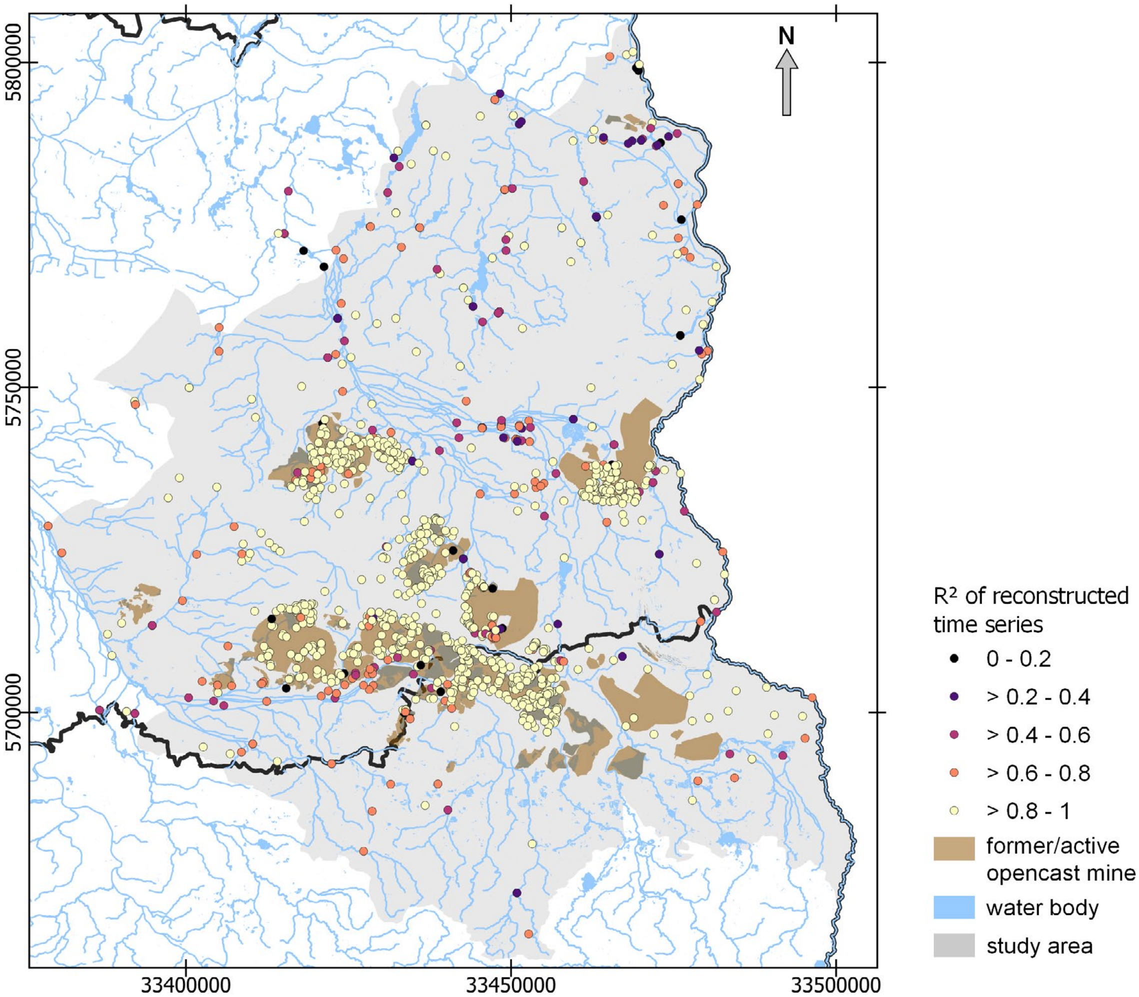


FIGURE 7 | Map of R^2 distribution for additional 1253 groundwater and discharge time series as an indicator of reconstruction accuracy.

The random forest regression models for spatial prediction of the dominant effects on groundwater and runoff dynamics (PC_g1 to PC_g4) exhibited moderate predictive performance, with R^2 values of 53%, 46%, 33% and 25% for PC_g1, PC_g2, PC_g3 and PC_g4, respectively (Table 2). The variable importance rankings highlighted that land-use classes, geological characteristics, and ET loadings were among the five most predictive covariates in predicting PC_g1 to PC_g4 and thus played a crucial role in determining their spatial distribution (Table 2).

Although the R^2 values indicate limited explanatory power, the models were still able to capture key trends in the spatial distribution of effect sizes. The MAE and RMSE metrics, while not exceptionally low, indicate that the predictions align with the general spatial patterns of the observed data. Despite the moderate accuracy, the spatial distribution of the loadings of PC_g1 to 4 reflects the patterns observed in the point-wise groundwater and discharge measurements. The area-wide maps of the loadings for PC_g1 to 4 can be found in the Appendix A.

4 | Discussion

This study aimed at the identification and quantification of spatio-temporal patterns of various effects on hydrological dynamics in Lusatia. Our findings demonstrate the effectiveness of this data-driven approach in determining the main influences on groundwater, discharge and ET across the massively anthropogenically shaped lowland region.

4.1 | Key Spatio-Temporal Patterns in Ground-Based Time Series

The high loadings on PC_g1, predominantly located near former and active open-cast mines, underscore the significant influence of mining activities on groundwater dynamics within the region. However, PC_g1, which typically represents the average dynamics of the dataset (Gottschalk 1985; Lischeid et al. 2021) is primarily defined by the strong spatial clustering of monitoring sites around mining sites. The uneven distribution of measurement stations, with 80% of sites located within 10 km of mining areas (which covers 50% of the study region), suggests a potential sampling bias. This imbalance may amplify the dominance of mining-related effects in PC_g1, overshadowing other regional influences.

This dominance reflects the spatially extensive impacts of mining, as groundwater lowering and rebound can propagate over large areas due to high permeability of the subsurface

TABLE 2 | Random forest model covariates and accuracy for spatial prediction of first four principal components. Covariates were selected based on their variable importance: Only the five most important covariates, as identified in at least one of 100 model training iterations, were included in the RF regression models. The number of iterations in which each covariate ranked among the five most important variables is indicated in brackets.

PC	Covariates for RF model	Model validation R^2	Model validation RMSE	Model validation MAE	Model testing R^2	Model testing RMSE	Model testing MAE
1	PC_e2 (98) and PC_e3 (100) of PCA from ET data, groundwater table depth (100), land use (2), lithology (100), precipitation (100)	0.54 [0.04]	0.42 [0.02]	0.29 [0.02]	0.53	0.44	0.32
2	PC_e1 (98), PC_e2 (100) and PC_e3 (3) of PCA from ET data, groundwater table depth (100), land use (96), lithology (3), precipitation (100)	0.51 [0.07]	0.20 [0.01]	0.14 [0.01]	0.46	0.21	0.15
3	PC_e1 (93), PC_e2 (100) and PC_e3 (65) of PCA from ET data, groundwater table depth (100), available water content in soil (6), land use (8), lithology (29), precipitation (100)	0.41 [0.04]	0.21 [0.01]	0.16 [0.01]	0.34	0.23	0.17
4	PC_e1 (100), PC_e2 (100) and PC_e3 (39) of PCA from ET data, groundwater table depth (100), available water content in soil (5), land use (26), lithology (30), precipitation (100)	0.38 [0.07]	0.16 [0.02]	0.11 [0.01]	0.25	0.18	0.13

layers in the region. The long spatial correlation lengths inherent in lowland hydrology, characterised by extensive groundwater connectivity, make these mining-related dynamics not only locally but also regionally influential. This underlines the importance of incorporating approaches that are capable of resolving such spatially extensive impacts, as emphasised in the introduction.

PC_g2 appears to capture seasonal variations, as indicated by its correlation with time series from sites situated more than 10 km away from mining areas. Unlike PC_g1, which is primarily associated with anthropogenic influences, PC_g2 likely reflects natural hydrological processes, particularly seasonal fluctuations in water balance driven by precipitation and ET (Longuevergne, Florsch, and Elsass 2007; Lischeid et al. 2021). This suggests that PC_g2 represents groundwater dynamics less affected by mining activities, offering insight into the region's background hydrology. Specifically, the spatial distribution of PC_g2 loadings, which are concentrated in areas away from direct mining impacts, highlights their role in delineating hydrological regions shaped by natural variability rather than human alteration. Such insights are crucial for distinguishing fundamental hydrological processes from the superimposed effects of mining and provide a key reference point for sustainable water management in lowland regions.

PC_g3 and PC_g4, while explaining a smaller portion of the variance, reveal more localised yet significant patterns. PC_g3 reflects delayed groundwater responses to mining, particularly in areas where groundwater levels have fluctuated as a result of decommissioning and controlled rebound efforts (Hangen-Brodersen, Strempel, and Grünewald 2005). The distinct temporal trend in PC_g3's time series, marked between 1995 and 2011, coincides with the closure of several mining operations a few years earlier in the western parts of the region, leading to a managed recovery in groundwater levels. This delayed response points to the spatial distance from active mining sites and the extended timeframe over which groundwater adjustments occur.

Loadings of PC_g4, on the other hand, are closely correlated with mean groundwater table depth, suggesting that it represents processes associated with groundwater recharge. The spatial pattern of loadings of PC_g4 captures the degree of attenuation and delay in the precipitation signal as it passes through the unsaturated zone, an effect well-documented in lowland catchments (Hohenbrink and Lischeid 2015; Lehr and Lischeid 2020). This effect, detected through PCA, emphasises the role of groundwater recharge in shaping the region's water dynamics, independent of mining activities.

Interestingly, while previous studies, such as Thomas et al. (2012), have found correlations between land cover and PCs in PCA of discharge data, such effects were not prominent in the first four PCs of this study. The absence of significant land cover correlations in PC_g 1 to 4 suggests that the overriding influence of mining activities likely masks subtle land use effects that may be present in groundwater and runoff dynamics. Such land use effects are clearly revealed by the PCA of the ET time series.

4.2 | Key Spatio-Temporal Patterns in ET Time Series

The PCA of ET time series highlighted the spatial and temporal heterogeneity in ET patterns across the region. The time series of PC_e1 exhibits a clear seasonal pattern, with peak values in June and July and the lowest values from December to January. PC_e1 thus represents the dominant seasonal cycle for most of the study area, capturing the annual rhythm of ET driven by temperature and vegetation growth cycles. This mean behaviour is consistent across the landscape but is modified by PC_e2 and PC_e3 at specific locations, which account for deviations such as amplified or reduced seasonality and potential phase shifts in ET dynamics. This pattern is consistent with the findings of Lei, Ren, and Bian (2016), who applied PCA to MODIS NDVI time series in a semiarid mining area and observed that the first PC captured the annual cycle of vegetation dynamics. Similarly, in our study, the spatial distribution of PC_e1 loadings highlights deviations from this seasonal pattern, particularly in areas affected by mining activity or with sparse vegetation. These areas exhibit lower ET values due to reduced water availability and vegetation cover over the past 30 years. Lei, Ren, and Bian (2016) also noted high values in areas with dense vegetation coverage, indicating a pronounced annual cycle, while areas with sparse vegetation had less distinct seasonal patterns. This comparison reinforces the strong influence of vegetation density and water availability on ET dynamics.

PC_e2 shows a strong correlation with land use, particularly related to soil water availability and the variability of ET across different land use types. Over the last 30 years, agricultural areas have experienced shifts in ET patterns, especially in response to climate variations. During the warmer and drier years (e.g., 2015, 2016, 2018–2020, 2022), ET in agricultural regions tends to peak earlier in the growing season (May/June), after which it drops sharply. In contrast, forested areas maintain higher ET during the summer months, due to increased water availability and sustained transpiration (Kleine et al. 2020; Landgraf et al. 2022). Notably, during the 1990s, agricultural areas exhibited higher ET earlier in the season compared to forests and water bodies, but this trend has reversed in more recent years. This suggests that the effect of recent climatic shifts disproportionately impacts agricultural areas.

The evidence from Trnka et al. (2011) also highlights that agricultural areas are experiencing increased vulnerability to climatic shifts, particularly with regard to ETET and water availability. Rising temperatures and increased drought stress have shortened the growing seasons and caused earlier collapses of ET in arable land compared to forests, which benefit from deeper root systems and higher soil water retention, allowing them to maintain more stable ET rates (Trnka et al. 2011; Teuling et al. 2019). Additionally, Renner and Hauffe (2024) emphasises that forest areas, due to deeper rooting depths and greater water retention capacities, demonstrate more resilience in ET behaviour under drought conditions compared to croplands. These findings corroborate the trend of arable land being more sensitive to the increasing ET demands of the atmosphere caused by recent climatic shifts.

PC_e2 also captures long-term changes in ET in regions affected by mining activities. The lowering of groundwater tables in and around active open-cast mining areas is reflected in the negative loadings of PC_e2, indicating a sustained decrease in ET. This suggests that long-term groundwater depletion significantly influences ET patterns.

PC_e3 highlights long-term trends in ET related to land use changes, particularly in areas affected by mining and restoration activities. Areas with active mining or deforestation show negative loadings on PC_e3 due to ongoing degradation of vegetation cover and reduced transpiration. Conversely, regions undergoing natural succession or restoration, such as former military training areas and afforestation sites, exhibit positive loadings on PC_e3, signifying an increasing trend in ET as vegetation recovers. This pattern aligns with findings from Lei, Ren, and Bian (2016), where vegetation dynamics, as indicated by NDVI, played a critical role in shaping ET patterns. The similarity between these studies highlights the close relationship between ET and vegetation characteristics, particularly NDVI, which is a crucial input parameter in the SSEBOp model used for ET estimation of Landsat data.

The use of remote sensing data for ET estimation further complements ground-based measurements, allowing for a more comprehensive analysis of water dynamics in lowland regions. This approach addresses the limitations of traditional monitoring methods in capturing both natural and anthropogenic processes driving ET changes over large areas.

The clear separation of ET patterns driven by water and vegetation dynamics underscores the effectiveness of PCA in capturing complex temporal and spatial variations in ET. Mascaro, Vivoni, and Méndez-Barroso (2015) similarly demonstrated that vegetation is one of the primary factors influencing the spatial distribution of ET. Nevertheless, their PCA relied on model-based ET time series, which may be affected by underlying model assumptions.

However, some limitations of ET analysis should be acknowledged. The derivation of ET time series from LST data, while physically based, operates at a global scale and may not fully account for regional or local factors influencing ET (Guerschman et al. 2022). Additionally, gaps in the time series due to cloud cover, as well as data limitations from Landsat 7 failure, introduce uncertainties in regions where temporal interpolation was necessary. Despite these limitations, our findings demonstrate that PCA applied to remote sensing-based ET time series can effectively detect changes in this substantial part of the water balance and thus supplementing ground-based hydrological monitoring with a more complete picture of the variables influencing regional water balance.

4.3 | Mapping of Primary Factors Influencing the Water Balance

The results indicate that the key factors influencing water balance in Lusatia can be effectively identified and spatially delineated using data from different water balance components and PCA. Despite the complexity of the landscape, characterised by significant anthropogenic influences such as mining, water withdrawals and river management, the RF models successfully explained essential portions of the variance in groundwater dynamics using relatively few and easily accessible input variables. This underscores the strength of data-driven approaches, which can capture key patterns even in highly modified environments (McPhee and Yeh 2008; Yu and Chu 2010).

The variable importance rankings in the RF models highlighted land-use classes, geological characteristics and ET loadings as the most predictive covariates for the spatial distribution of the PC loadings. This finding emphasises the critical role both anthropogenic factors (e.g., land use change, mining activity) and natural processes (e.g., vegetation dynamics, geological conditions) play in shaping groundwater dynamics. For instance, land-use classes likely capture both anthropogenic influences, such as changes in agricultural or industrial activity and natural processes, such as vegetation growth and water use. Similarly, geological characteristics describe the permeability and storage capacity of groundwater, while ET loadings indicate water availability through the impact of ET processes.

The moderate predictive performance of the RF models, as indicated by relatively modest R^2 values and error metrics (MAE and RMSE), reflects the inherent challenges of modelling complex hydrological systems using statistical approaches that lack physically based assumptions. However, these models still captured general spatial trends and major hydrological drivers, despite limitations like the uneven spatial distribution of monitoring sites, particularly near mining areas. This uneven sampling may have impacted the model's ability to generalise predictions across regions with less dense monitoring data, yet the results still align well with the observed effects from the ground-based dataset.

The maps generated from RF predictions offer valuable insights, particularly for areas where monitoring data is sparse or unavailable. While not fully accurate, these maps provide a useful representation of dominant hydrological processes serving as a preliminary tool for identifying regions where groundwater dynamics may be influenced by specific factors, such as land use or geology.

In conclusion, while the RF-based approach has limitations due to its moderate accuracy and lack of physically based assumptions, it nonetheless provides meaningful insights into the spatial patterns of groundwater dynamics in Lusatia. This method represents a useful complement to traditional hydrological modelling, particularly in regions with limited data availability. We see great potential in a systematic iterative approach merging the proposed statistical approach with physically based models to enhance the predictive accuracy and better capture the complexity of the hydrological processes at play.

5 | Conclusions

This study demonstrates the effectiveness of a data-driven approach, using PCA to answer the three main objectives of our study. First, we examined the key spatio-temporal patterns in ground-based time series of groundwater and discharge and attempted to distinguish between anthropogenic and natural effects. Our analysis revealed significant differences in the patterns, with the results highlighting the dominance of anthropogenic influences, particularly mining activities and the seasonal fluctuations caused by natural processes such as precipitation and ET. PCA allowed us to effectively differentiate these effects and gain a clearer understanding of the groundwater dynamics in the region. Second, we examined the main spatio-temporal patterns in ET time series and assessed how they correlate with specific land use types and climate factors. Our results showed pronounced ET dynamics with notable correlations between land-use types—such as agricultural land, forests and water bodies—and seasonal variations in ET. Finally, we investigated how the integration of groundwater, surface water and ET time series data can map the primary factors influencing the water balance in both spatial and temporal terms. By combining these datasets with additional environmental parameters, we mapped the most important factors and their spatio-temporal patterns across the entire study area.

Our findings demonstrate the value of PCA as a powerful tool for analysing complex hydrological interactions and offering valuable insights for spatio-temporal hydrological assessments in regions with limited data. The findings provide a robust basis for further research and can serve as a preliminary tool for guiding future hydrological investigations and informing water management strategies. Building on this foundation, the approach can be refined through the use of more granular datasets or alternative modelling techniques to enhance the predictive accuracy.

Acknowledgements

Groundwater head, surface water level and runoff data have been kindly provided by the Landesamt für Umwelt (Brandenburg State Office for the Environment [LfU]), the Landesamt für Umwelt, Landwirtschaft und Geologie (Saxony State Office for the Environment, Agriculture and Geology [LfULG]), the Bundesamt für Gewässerkunde (Federal Institute of Hydrology [BfG]) and the Lausitzer and Mitteldeutsche Bergbau-Verwaltungsgesellschaft mbH (LMBV). These contributions are highly appreciated. This study has been funded by the German Federal Ministry of Education and Research (BMBF, Grant number 03WIR3020B).

Data Availability Statement

Satellite images of Landsat 5, 7 and 8 and Copernicus data are available for free use (https://espa.cr.usgs.gov/; https://land.copernicus.eu/en). Precipitation data from German Weather Service are available via https://www.dwd.de/EN/. Measurements of groundwater and surface water level and runoff can be requested from the State Offices. Data on soil property and geology are also available for free use (https://inspire-geoportal.ec.europa.eu/).

References

Barthel, R., and S. Banzhaf. 2016. "Groundwater and Surface Water Interaction at the Regional-Scale—A Review With Focus on Regional Integrated Models." *Water Resources Management* 30, no. 1: 1–32. https://doi.org/10.1007/s11269-015-1163-z.

Barthel, R., E. Haaf, M. Giese, M. Nygren, B. Heudorfer, and K. Stahl. 2021. "Similarity-Based Approaches in Hydrogeology: Proposal of a New Concept for Data-Scarce Groundwater Resource Characterization and Prediction." *Hydrogeology Journal* 29, no. 5: 1693–1709. https://doi.org/10.1007/s10040-021-02358-4.

Biemelt, D., A. Schapp, and U. Grünewald. 2011. "Hydrological Observation and Modelling Relationship for the Determination of Water Budget in Lusatian Post-Mining Landscape." *Physics and Chemistry of the Earth, Parts A/B/C* 36, no. 1–4: 3–18. https://doi.org/10.1016/j.pce.2010.07.015.

Blöschl, G., and M. Sivapalan. 1995. "Scale Issues in Hydrological Modelling: A Review." *Hydrological Processes* 9, no. 3–4: 251–290. https://doi.org/10.1002/hyp.3360090305.

Bretherton, C. S., C. Smith, and J. M. Wallace. 1992. "An Intercomparison of Methods for Finding Coupled Patterns in Climate Data." *Journal of Climate* 5, no. 6: 541–560. https://doi.org/10.1175/1520-0442(1992)005<0541:AIOMFF>2.0.CO;2.

Condon, L. E., S. Kollet, M. F. P. Bierkens, et al. 2021. "Global Groundwater Modeling and Monitoring: Opportunities and Challenges." *Water Resources Research* 57, no. 12: e2020WR029500. https://doi.org/10.1029/2020WR029500.

Deutscher Wetterdienst (DWD). 2023a. "Lufttemperatur: vieljährige Mittelwerte 1991–2020." Last modified August 16, 2023. https://www.dwd.de/DE/leistungen/klimadatendeutschland/vielj_mittelwerte.html.

Deutscher Wetterdienst (DWD). 2023b. "Niederschlag: vieljährige Mittelwerte 1991–2020." Last modified July 26, 2023. https://www.dwd.de/DE/leistungen/klimadatendeutschland/vielj_mittelwerte.html.

Deutscher Wetterdienst (DWD). 2024. "Abgeleitete Parameter an Stationsstandorten. potentielle Grasreferenzverdunstung nach FAO." Last modified March 22, 2024. https://opendata.dwd.de/climate_environment/CDC/derived_germany/soil/multi_annual/derived_germany_soil_multi_annual_v2_mean_91-20_VPGFAO.txt.

DWD Climate Data Center (CDC). 2021. "Raster der vieljährigen Mittel der Niederschlagshöhe für Deutschland." Version v21.3. https://dwd-geoportal.de/products/GRD_DEU_P30Y_RR_P1M/.

European Environment Agency. 2019a. "CORINE Land Cover 2018 (Vector), Europe, 6-Yearly—Version 2020_20u1, May 2020."

European Environment Agency. 2019b. "CORINE Land Cover Change 1990–2000 (Vector), Europe, 6-Yearly—Version 2020_20u1, May 2020."

European Environment Agency. 2019c. "CORINE Land Cover Change 2000–2006 (Vector), Europe, 6-Yearly—Version 2020_20u1, May 2020."

European Environment Agency. 2019d. "CORINE Land Cover Change 2006–2012 (Vector), Europe, 6-Yearly—Version 2020_20u1, May 2020."

European Environment Agency. 2019e. "CORINE Land Cover Change 2012–2018 (Vector), Europe, 6-Yearly—Version 2020_20u1, May 2020."

Gädeke, A., I. Pohle, H. Koch, and U. Grünewald. 2017. "Trend Analysis for Integrated Regional Climate Change Impact Assessments in the Lusatian River Catchments (North-Eastern Germany)." *Regional Environmental Change* 17, no. 6: 1751–1762. https://doi.org/10.1007/s10113-017-1138-0.

Gerwin, W., T. Raab, K. Birkhofer, et al. 2023. "Perspectives of Lignite Post-Mining Landscapes Under Changing Environmental Conditions: What Can We Learn From a Comparison Between the Rhenish and Lusatian Region in Germany?" *Environmental Sciences Europe* 35, no. 1: 36. https://doi.org/10.1186/s12302-023-00738-z.

Giese, M., E. Haaf, B. Heudorfer, and R. Barthel. 2020. "Comparative Hydrogeology—Reference Analysis of Groundwater Dynamics From Neighbouring Observation Wells." *Hydrological Sciences Journal* 65, no. 10: 1685–1706. https://doi.org/10.1080/02626667.2020.1762888.

Gottschalk, L. 1985. "Hydrological Regionalization of Sweden." *Hydrological Sciences Journal* 30, no. 1: 65–83. https://doi.org/10.1080/02626668509490972.

Guerschman, J. P., T. R. McVicar, J. Vleeshower, T. G. van Niel, J. L. Peña-Arancibia, and Y. Chen. 2022. "Estimating Actual Evapotranspiration at Field-To-Continent Scales by Calibrating the CMRSET Algorithm With MODIS, VIIRS, Landsat and Sentinel-2 Data." *Journal of Hydrology* 605: 127318. https://doi.org/10.1016/j.jhydrol.2021.127318.

Haaf, E., M. Giese, T. Reimann, and R. Barthel. 2023. "Data-Driven Estimation of Groundwater Level Time-Series at Unmonitored Sites Using Comparative Regional Analysis." *Water Resources Research* 59, no. 7: e2022WR033470. https://doi.org/10.1029/2022WR033470.

Hangen-Brodersen, C., P. Strempel, and U. Grünewald. 2005. "Characteristics of Catchments Disturbed by Lignite Mining—Case Study of Schlabendorf/Seese (Germany)." *Ecological Engineering* 24, no. 1–2: 37–48. https://doi.org/10.1016/j.ecoleng.2004.12.005.

Heudorfer, B., E. Haaf, K. Stahl, and R. Barthel. 2019. "Index-Based Characterization and Quantification of Groundwater Dynamics." *Water Resources Research* 55, no. 7: 5575–5592. https://doi.org/10.1029/2018WR024418.

Hofmann, G., and U. Pommer. 2005. "Potentielle natürliche Vegetation von Brandenburg und Berlin." Eberswalde (Eberswalder Forstliche Schriftenreihe, XXIV).

Hohenbrink, T. L., and G. Lischeid. 2015. "Does Textural Heterogeneity Matter? Quantifying Transformation of Hydrological Signals in Soils." *Journal of Hydrology* 523: 725–738. https://doi.org/10.1016/j.jhydrol.2015.02.009.

Kleine, L., D. Tetzlaff, A. Smith, H. Wang, and C. Soulsby. 2020. "Using Water Stable Isotopes to Understand Evaporation, Moisture Stress, and Re-Wetting in Catchment Forest and Grassland Soils of the Summer Drought of 2018." *Hydrology and Earth System Sciences* 24, no. 7: 3737–3752. https://doi.org/10.5194/hess-24-3737-2020.

Krümmelbein, J., O. Bens, T. Raab, and M. Anne Naeth. 2012. "A History of Lignite Coal Mining and Reclamation Practices in Lusatia, Eastern Germany." *Canadian Journal of Soil Science* 92, no. 1: 53–66. https://doi.org/10.4141/cjss2010-063.

Kuhn, M. 2008. "Building Predictive Models in R Using the Caret Package." *Journal of Statistical Software* 28, no. 5: 1–26. https://doi.org/10.18637/jss.v028.i05.

Kumar, P., S. E. Debele, J. Sahani, et al. 2021b. "Nature-Based Solutions Efficiency Evaluation Against Natural Hazards: Modelling Methods, Advantages and Limitations." *Science of the Total Environment* 784: 147058. https://doi.org/10.1016/j.scitotenv.2021.147058.

Kumar, P., S. E. Debele, J. Sahani, et al. 2021a. "An Overview of Monitoring Methods for Assessing the Performance of Nature-Based Solutions Against Natural Hazards." *Earth-Science Reviews* 217: 103603. https://doi.org/10.1016/j.earscirev.2021.103603.

Landgraf, J., D. Tetzlaff, S. Wu, J. Freymüller, and C. Soulsby. 2022. "Using Stable Water Isotopes to Understand Ecohydrological Partitioning Under Contrasting Land Uses in a Drought-Sensitive Rural, Lowland Catchment." *Hydrological Processes* 36, no. 12: e14779. https://doi.org/10.1002/hyp.14779.

Lehr, C., and G. Lischeid. 2020. "Efficient Screening of Groundwater Head Monitoring Data for Anthropogenic Effects and Measurement Errors." *Hydrology and Earth System Sciences* 24, no. 2: 501–513. https://doi.org/10.5194/hess-24-501-2020.

Lei, S., L. Ren, and Z. Bian. 2016. "Time–Space Characterization of Vegetation in a Semiarid Mining Area Using Empirical Orthogonal Function Decomposition of MODIS NDVI Time Series." *Environmental Earth Sciences* 75, no. 6: 516. https://doi.org/10.1007/s12665-015-5122-z.

Lischeid, G., R. Dannowski, K. Kaiser, G. Nützmann, J. Steidl, and P. Stüve. 2021. "Inconsistent Hydrological Trends Do Not Necessarily Imply Spatially Heterogeneous Drivers." *Journal of Hydrology* 596: 126096. https://doi.org/10.1016/j.jhydrol.2021.126096.

Longuevergne, L., N. Florsch, and P. Elsass. 2007. "Extracting Coherent Regional Information From Local Measurements With Karhunen-Loève Transform: Case Study of an Alluvial Aquifer (Rhine Valley, France and Germany)." *Water Resources Research* 43, no. 4: W04430. https://doi.org/10.1029/2006WR005000.

Mascaro, G., E. R. Vivoni, and L. A. Méndez-Barroso. 2015. "Hyperresolution Hydrologic Modeling in a Regional Watershed and Its Interpretation Using Empirical Orthogonal Functions." *Advances in Water Resources* 83: 190–206. https://doi.org/10.1016/j.advwatres.2015.05.023.

McPhee, J., and W. W.-G. Yeh. 2008. "Groundwater Management Using Model Reduction via Empirical Orthogonal Functions." *Journal of Water Resources Planning and Management* 134, no. 2: 161–170. https://doi.org/10.1061/(ASCE)0733-9496(2008)134:2(161).

Meggiorin, M., P. Bullo, V. Accoto, G. Passadore, A. Sottani, and A. Rinaldo. 2022. "Applying the Principal Component Analysis for a

Deeper Understanding of the Groundwater System: Case Study of the Bacchiglione Basin (Veneto, Italy).” *Acque Sotterranee-Italian Journal of Groundwater* 11, no. 2: 7–17. https://doi.org/10.7343/as-2022-573.

Millard, K., and M. Richardson. 2015. “On the Importance of Training Data Sample Selection in Random Forest Image Classification: A Case Study in Peatland Ecosystem Mapping.” *Remote Sensing* 7, no. 7: 8489–8515. https://doi.org/10.3390/rs70708489.

Page, R. M., G. Lischeid, J. Epting, and P. Huggenberger. 2012. “Principal Component Analysis of Time Series for Identifying Indicator Variables for Riverine Groundwater Extraction Management.” *Journal of Hydrology* 432–433: 137–144. https://doi.org/10.1016/j.jhydrol.2012.02.025.

Pebesma, E. 2018. “Simple Features for R: Standardized Support for Spatial Vector Data.” *R Journal* 10, no. 1: 439–446. https://doi.org/10.32614/RJ-2018-009.

Pebesma, E., and R. Bivand. 2023. *Spatial Data Science: With Applications in R*, 352. New York: Chapman and Hall/CRC. https://doi.org/10.1201/9780429459016.

Petrakis, R. E., L. M. Norman, M. L. Villarreal, et al. 2024. “An Ensemble Mean Method for Remote Sensing of Actual Evapotranspiration to Estimate Water Budget Response Across a Restoration Landscape.” *Remote Sensing* 16, no. 12: 2122. https://doi.org/10.3390/rs16122122.

R Core Team. 2024. *R. A Language and Anvironment for Statistical Computing. Version 4.2.3*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Renner, M., and C. Hauffe. 2024. “Impacts of Climate and Land Surface Change on Catchment Evapotranspiration and Runoff From 1951 to 2020 in Saxony, Germany.” *Hydrology and Earth System Sciences* 28, no. 13: 2849–2869. https://doi.org/10.5194/hess-28-2849-2024.

Reyer, C., J. Bachinger, R. Bloch, et al. 2012. “Climate Change Adaptation and Sustainable Regional Development: A Case Study for the Federal State of Brandenburg, Germany.” *Regional Environmental Change* 12, no. 3: 523–542. https://doi.org/10.1007/s10113-011-0269-y.

Rinderer, M., H. J. van Meerveld, and B. L. McGlynn. 2019. “From Points to Patterns: Using Groundwater Time Series Clustering to Investigate Subsurface Hydrological Connectivity and Runoff Source Area Dynamics.” *Water Resources Research* 55, no. 7: 5784–5806. https://doi.org/10.1029/2018WR023886.

Senay, G. B. 2018. “Satellite Psychrometric Formulation of the Operational Simplified Surface Energy Balance (SSEBop) Model for Quantifying and Mapping Evapotranspiration.” *Applied Engineering in Agriculture* 34, no. 3: 555–566. https://doi.org/10.13031/aea.12614.

Senay, G. B., M. K. Friedrichs, C. Morton, et al. 2022. “Mapping Actual Evapotranspiration Using Landsat for the Conterminous United States: Google Earth Engine Implementation and Assessment of the SSEBop Model.” *Remote Sensing of Environment* 275: 113011. https://doi.org/10.1016/j.rse.2022.113011.

Senay, G. B., G. E. L. Parrish, M. Schauer, et al. 2023. “Improving the Operational Simplified Surface Energy Balance Evapotranspiration Model Using the Forcing and Normalizing Operation.” *Remote Sensing* 15, no. 1: 260. https://doi.org/10.3390/rs15010260.

Statistik der Kohlewirtschaft. 2024. “Landinanspruchnahme, Rekultivierung. Betriebsflächen und wieder nutzbar gemachte Flächen im Braunkohlenbergbau in Deutschland. Stand: Ende Dezember 2023. Statistik der Kohlewirtschaft e.V.” Last modified March 11, 2024. https://kohlenstatistik.de/downloads/braunkohle/.

Teuling, A. J., E. A. G. de Badts, F. A. Jansen, et al. 2019. “Climate Change, Reforestation/Afforestation, and Urbanization Impacts on Evapotranspiration and Streamflow in Europe.” *Hydrological Earth System Science* 23: 3631–3652. https://doi.org/10.5194/hess-23-3631-2019.

Thomas, B., G. Lischeid, J. Steidl, and R. Dannowski. 2012. “Regional Catchment Classification With Respect to Low Flow Risk in a Pleistocene Landscape.” *Journal of Hydrology* 475: 392–402. https://doi.org/10.1016/j.jhydrol.2012.10.020.

Trnka, M., J. Eivind Olesen, K. C. Kersebaum, et al. 2011. “Agroclimatic Conditions in Europe Under Climate Change.” *Global Change Biology* 17, no. 7: 2298–2318. https://doi.org/10.1111/j.1365-2486.2011.02396.x.

Uhlmann, W., K. Zimmermann, T. Claus, C. Gerstgraser, and S. Giebler. 2020. “Erarbeitung eines strategischen Hintergrundpapiers zu den bergbaubedingten Stoffeinträgen in den Flusseinzugsgebieten Spree und Schwarze Elster.” ElsterTeil 2: Zustandsanalyse und Handlungsschwerpunkte. Dresden/Cottbus.

Uhlmann, W., K. Zimmermann, M. Kaltofen, C. Gerstgraser, F. Grosser, and C. Schützel. 2023. “Wasserwirtschaftliche Folgen des Braunkohleausstiegs in der Lausitz. Abschlussbericht. Hg. v.” Umweltbundesamt (UBA). Dessau-Roßlau (Texte, 90/2023).

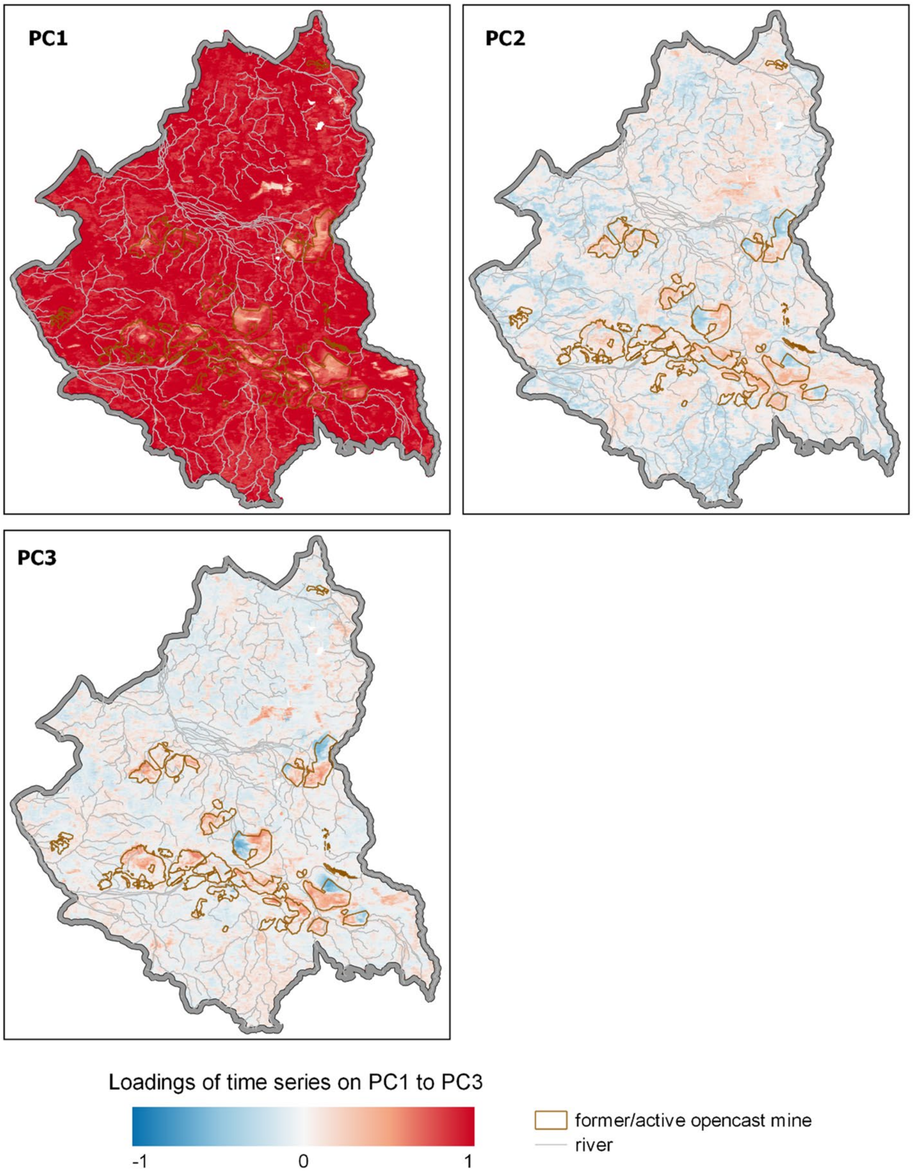
Wickham, H., R. François, L. Henry, K. Müller, and D. Vaughan. 2023. “dplyr: A Grammar of Data Manipulation. R Package Version 1.1.2.” https://CRAN.R-project.org/package=dplyr.

Yu, H.-L., and H.-J. Chu. 2010. “Understanding Space–Time Patterns of Groundwater System by Empirical Orthogonal Functions: A Case Study in the Choshui River Alluvial Fan, Taiwan.” *Journal of Hydrology* 381, no. 3–4: 239–247. https://doi.org/10.1016/j.jhydrol.2009.11.046.

Yue, W., K. Meng, K. Hou, R. Zuo, B.-T. Zhang, and G. Wang. 2020. “Evaluating Climate and Irrigation Effects on Spatiotemporal Variabilities of Regional Groundwater in an Arid Area Using EOFs.” *Science of the Total Environment* 709: 136147. https://doi.org/10.1016/j.scitotenv.2019.136147.

Appendix A

Spatial patterns of loadings of PC_e1 to PC_e4 of Landsat-ET time series.



Spatial patterns of loadings of PC_g1 to PC_g4 based on RF model of ground-based time series.