Received: 22 February 2023 | Revised: 24 August 2023 | Accepted: 28 August 2023

DOI: 10.1002/hyp.14991

RESEARCH ARTICLE

WILEY

Predicting stream water temperature with artificial neural networks based on open-access data

Konstantina Drainas[1] | Lisa Kaule[2] | Stefanie Mohr[3] | Bhumika Uniyal[4] | Romy Wild[1] | Juergen Geist[1]

[1]Aquatic Systems Biology, TUM School of Life Sciences, Technical University of Munich, Freising, Germany

[2]Department of Hydrology, Bayreuth Center of Ecology and Environmental Research (BayCEER), University of Bayreuth, Bayreuth, Germany

[3]Foundations of Software Reliability and Theoretical Computer Science, TUM School of Computation, Information and Technology, Garching, Germany

[4]Professorship of Ecological Services, Bayreuth Center of Ecology and Environmental Research (BayCEER), University of Bayreuth, Bayreuth, Germany

Correspondence
Juergen Geist, Aquatic Systems Biology, TUM School of Life Sciences, Technical University of Munich, Freising, Germany.
Email: geist@tum.de

Funding information
Bayerisches Staatsministerium für Wissenschaft und Kunst; Deutsche Forschungsgemeinschaft, Grant/Award Number: GRK 2428

Abstract

Predictions of stream water temperature are an important tool for assessing potential impacts of climate warming on aquatic ecosystems and for prioritizing targeted adaptation and mitigation measures. Since predictions require reliable baseline data, we assessed whether open-access data can serve as a suitable resource for accurate and reliable water temperature prediction using artificial neural networks (ANNs). For this purpose, we trained and tested ANNs in 16 small ($\leq 1 \frac{m^3}{s}$) headwater streams of major types located in Bavaria, Germany. Between four and eight different combinations of input parameters were trained and tested for each stream ANN, based on data availability. These were air temperature (mean, minimum and maximum), day of the year, discharge, water level and sunshine duration per day. We found that the input combination with the highest accuracy (lowest RMSE) was stream-specific, suggesting that the optimal input combination cannot be generalized across streams. Using a reasonable, but random, input combination resulted in an increase in error (RMSE) of up to >100% compared to the stream-specific optimal combination. Hence, we conclude that the accuracy of water temperature prediction strongly depends on the availability of open-access input data. We also found that environmental parameters such as hydrological characteristics and the proportion of land use in the 5 m riparian strip and the entire catchment are important drivers, affecting the accuracy and reliability of ANNs. ANNs' prediction accuracy was strongly negatively related to river length, total catchment area and water level. High proportions of semi-natural and forested land cover correlated with a higher accuracy, while open-canopy land use types such as grassland were negatively associated with ANN accuracy. In conclusion, open-access data were found to be suitable for accurate and reliable predictions of water temperature using ANNs. However, we recommend incorporating stream-specific environmental information and tailor the combination of input parameters to individual streams in order to obtain optimal results.

KEYWORDS
artificial neural networks, climate change, machine learning, open-access data, prediction, stream habitat quality, thermal stress, water temperature modelling

1 | INTRODUCTION

With rising atmospheric temperatures, climate change affects stream water temperatures due to the well-established relationship between air and water temperature (Crisp & Howson, 1982; Kothandaraman & Evans, 1972; Mohseni & Stefan, 1999; Webb et al., 2003). This is of particular relevance in small headwater streams, where the relatively low mean stream depth is highly influenced by surface energy fluxes (Leach et al., 2023), which are correlated with air temperature (e.g. solar radiation). Moreover, headwater areas govern processes further downstream and their coldwater spots provide important refugia for coldwater-dependent species in light of climatic change (Kuhn et al., 2021). Stream water temperature is naturally regulated by various drivers: meteorological (air temperature and net radiation), hydro(geo)logical (discharge and groundwater inflow), hydromorphological (stream width and depth), and vegetational, the latter determining shading and evapotranspiration. There are a number of anthropogenic activities that influence variables such as discharge and flow variation, the proportion of surfaces rendered impervious by urbanization, changes in ice cover, and thermal pollution, which can additionally affect the water's thermal properties and natural temperature regimes on a large spatial scale (Caldwell et al., 2015; Nelson & Palmer, 2007). As one of the most common forms of anthropogenic disturbance to ecosystems, land use plays a key role in water temperature regulation, for example due to the partitioning of precipitation into infiltration and surface runoff, which affects water regimes on a (sub-)catchment scale.

Since temperature is the most crucial determinant of abiotic and biotic processes, further anticipated changes in stream water temperature due to global warming and other human impacts are expected to have substantial effects on aquatic ecosystems (Smith, 1981). This not only includes a decreased saturation concentration of oxygen triggered by global warming (Piatka et al., 2021), but also changes in viscosity, vapour pressure, density, and surface tension (Caissie, 2006). Additionally, temperature controls a wide range of biological processes, such as the decomposition rate of organic matter, species composition in aquatic communities, biotic interactions, and energy transfer in aquatic food webs (Woodward et al., 2010). The rapid pace of global warming (IPCC, 2022) creates a need for more detailed predictions of future water temperatures in streams. These are urgently required to enable an assessment of the potential impacts of climate warming on the abiotic stream environment and the consequences for biological communities. An understanding of this is also key to targeting and prioritizing mitigation and adaptation measures.

The importance of predicting stream water temperatures is reflected by the variety of approaches that have already been tested. As stated in Rabi et al. (2015), water temperature prediction models can generally be divided into two major categories: deterministic and statistical. Statistical models are in turn differentiated into parametric and non-parametric ones (for definitions, see e.g. Rabi et al. (2015) or Benyahya et al. (2007)). The availability of data for deterministic models, such as SHADE (Chen et al., 1998) or CEQUEAU (St-Hilaire et al., 2000), is problematic, as many variables are required for catchment and thermal representations, along with complete time series for discharge and meteorological parameters. While parametric statistical models have much lower data requirements and are simple to use, their structure is specified from the start and hence not flexibly adjustable to the data (Benyahya et al., 2007). This limitation can lead to incorrect water temperature predictions when using linear regression, a technique often applied to describe the relationship between air and water temperature (Ahmadi-Nedushan et al., 2007; Crisp & Howson, 1982; Harvey et al., 2011; Krider et al., 2013; Rabi et al., 2015; Smith, 1981; Webb et al., 2003). At elevated and low air temperatures, physical effects lead to non-linearity (Mohseni & Stefan, 1999), which is beyond the limits of linear regression analysis.

In attempting to deal with the above challenges, the non-parametric statistical approach of Artificial Neural Networks (ANNs) is increasingly popular and has displayed equal or even higher accuracy (as evident from RMSE) than the majority of deterministic and parametric statistical models (Chenard & Caissie, 2008; Feigl et al., 2021; Hadzima-Nyarko et al., 2014; Pilgrim et al., 1998; Piotrowski et al., 2015; Rabi et al., 2015; Zhu et al., 2019,b,c). To the best of our knowledge, the smallest and hence best RMSE values reported for water temperature prediction using ANNs ranged between 0.46°C (Zhu, Heddam, Nyarko, et al., 2019) and 1.58°C (Hadzima-Nyarko et al., 2014). In the following, we refer to this "state of the art" range as "sota-range".

Besides performance, a major benefit of using ANNs compared to deterministic models are the lower data requirements. While deterministic models require large amounts of data for predictions of water temperature, ANNs have already displayed good results with comparably limited information. It is currently unknown which input parameters produce optimal results, and so their selection varies in different studies. While air temperature is a key input parameter, its format varies greatly, as do the additional input parameters, particularly those concerning the temporal resolution of the data. Several studies used only daily mean air temperatures or once-a-day-measurements (Graf et al., 2019; Hadzima-Nyarko et al., 2014; Qiu et al., 2020; Rabi et al., 2015; Zhu et al., 2019,b), while others are based on daily mean, minimum and maximum air temperatures (Chenard & Caissie, 2008; Feigl et al., 2021; Piotrowski et al., 2015). Most studies used discharge or water level (Chenard & Caissie, 2008; Feigl et al., 2021; Qiu et al., 2020; Zhu et al., 2019,b,c) and/or the day of the year as an additional input (Chenard & Caissie, 2008; Feigl et al., 2021; Hadzima-Nyarko et al., 2014; Qiu et al., 2020; Zhu et al., 2019,b,c), while only one study additionally used global radiation (Feigl et al., 2021) and one the declination of the sun (Piotrowski et al., 2015).

Various measures can be obtained to determine the quality of a prediction, the most prominent one being accuracy. Accuracy is an indicator of how exactly an ANN predicts the output in the context of training and testing. However, climate change and natural variability involve data variations that ANNs might not be sufficiently capable of learning, since data obtained for training and testing cannot be used to represent future climatic developments. It is therefore not sufficient to rely solely on accuracy to determine the suitability of an ANN for its task. For classification networks, there are several methods available that provide more insight into the behaviour of ANNs (for examples, see Bach et al., 2015; Baehrens et al., 2010; Erhan et al., 2009; Huang et al., 2020; Simonyan et al., 2013; and Sundararajan et al., 2017). However, the prediction of water temperature is not a classification but a regression problem. In the field of regression problems, Mohr et al. (2021), to

the best of our knowledge, were the first to develop a methodology, able to give insight into the behaviour of regression models and to measure their behaviour not only on the basis of accuracy calculations but also as a means of examining reliability. Consequently, we included these methods in our study to enable a more holistic picture of water temperature ANN performance.

While determining accuracy and reliability is important for understanding how much trust can be placed in a prediction, these measures do not fully explain the disturbances found in the predictions. Variations in environmental conditions and the land use form surrounding streams are highly relevant for explaining the behaviour of ANNs, how they are influenced by environmental factors, and which conditions allow for a reasonable use of ANNs for predicting water temperature.

Hence, in this study, we address whether it is possible to train accurate and reliable ANNs based on open-access data for small ($\leq 1\frac{m^3}{s}$) headwater streams in Bavaria, Germany. Additionally, we address whether an optimal combination of input parameters exists and whether these combinations are unique for each stream or can be generalized across streams. To confine the range of environmental conditions in which ANNs can be optimally applied, we studied how environmental parameters such as stream length, hydrological characteristics and proportion of land use types affect ANN accuracy. We hypothesized that open-access data suffices to predict water temperature in small headwater streams with an RMSE in the sota-range. We also hypothesized that the accuracy and reliability of ANNs are influenced by both the input combinations and the environmental parameters of the streams, which would make the optimal input combination stream-specific.

2 | MATERIALS AND METHODS

2.1 | Study sites

For this study, we investigated 16 streams in major eco-regions of different geological origins throughout Bavaria, Germany. Figure 1 shows the locations of the gauging stations by Gewaesserkundlicher Dienst Bayern (abbr. GkD) for each stream. Figures 2 and 3 depict water temperature time series that were available for each of the 16 gauging stations. We selected streams with a mean annual discharge of $\leq 1\frac{m^3}{s}$, based on open-access data from GkD. Using this criterion, we were able to focus on headwater streams, which are of special interest since they also govern processes further downstream.

2.2 | Measures of model performance

To assess the ANNs' performance, we used three different accuracy metrics as described in the following and three newly developed reliability metrics from Mohr et al. (2021) as described in Appendix A.3. Our aim was to prioritize the used accuracy metrics according to their expressive power regarding the reliability of ANNs. Therefore, we conducted correlation analysis (see description in Section 2.6.2) to identify connections between accuracy and reliability metrics. In the following, we describe the three used accuracy metrics and define them in Formulas 1, 2, and 3.

RMSE: According to Moriasi et al. (2007), the root mean square error (for RMSE, see Formula 1) is an error index commonly used in the context of model evaluation. A value of 0 indicates a perfect fit.

We chose this metric since it is regularly used in the context of water temperature predictions with ANNs and is intuitively well understandable.

R: The Pearson's product–moment correlation coefficient (for R, see Formula 2) describes the degree of collinearity between predicted and observed data (Rabi et al., 2015). It ranges from -1 to 1, where 0 indicates no linear relationship and -1 or 1 indicate a linear relationship. In this study, we aimed for a positive correlation between the observed and predicted values that is, values close to $+1$.

As the RMSE, this metric is regularly used in the context of water temperature prediction with ANNs but in the contrary does not show the mean error, but the degree of collinearity.

PBIAS: According to Gupta et al. (1999) as cited in Moriasi et al. (2007), the Percent bias (PBIAS, see Formula 3) shows whether the predictions are, on average, over- or underestimated. A value of 0 indicates a perfect fit, while positive values indicate underestimation and negative ones indicate overestimation.

This metric is uncommon in the field of water temperature prediction with ANNs but opens up a new perspective, since the direction of prediction inaccuracies (over- or underestimation) is displayed. On the contrary, the other two metrics concentrate, in general, on the amount of difference between observation and prediction.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(O_i - P_i)^2}, \tag{1}$$

$$R = \frac{\sum_{i=1}^{n}(O_i - \overline{O})\ (P_i - \overline{P})}{\sqrt{\sum_{i=1}^{n}(O_i - \overline{O})^2 \sum_{i=1}^{n}(P_i - \overline{P})^2}}, \tag{2}$$

$$\mathrm{PBIAS} = \frac{\sum_{i=1}^{n}(O_i - P_i) \times 100}{\sum_{i=1}^{n}(O_i)}. \tag{3}$$

To define the evaluation metrics, we followed the notation by Rabi et al. (2015), where P_i is the ith predicted water temperature value, O_i is the ith observed water temperature value, $\overline{P}$ is the average of P_i, $\overline{O}$ is the average of O_i and n is the size of the dataset.

2.3 | Input

The data basis for the ANNs consisted of open-access data supplied by the GkD and Germany's National Meteorological Service (DWD). The data used for each stream consisted of the daily mean water temperatures (°C), daily discharges (Q, m^3/s) and daily mean water levels

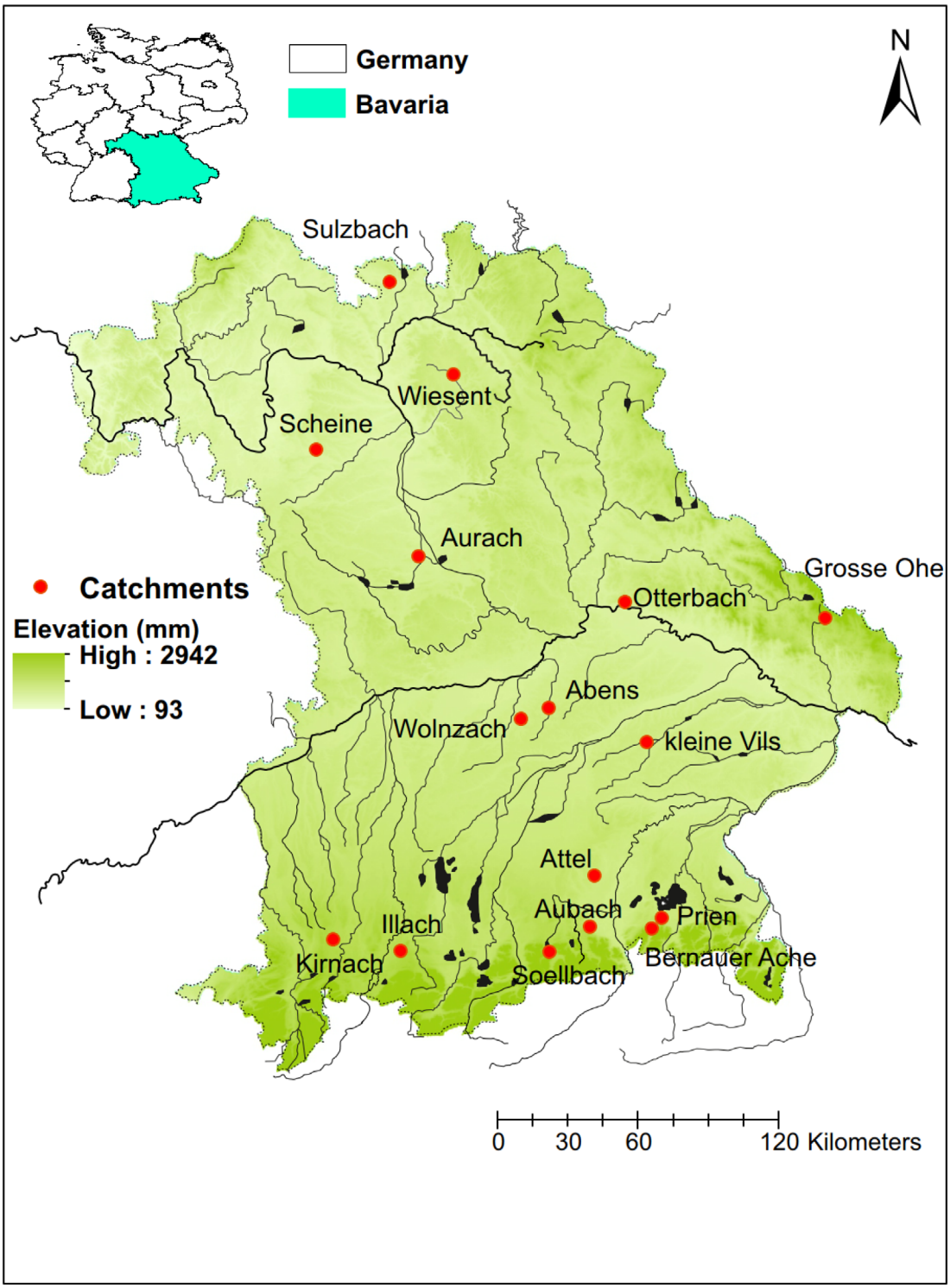

FIGURE 1 Location of GkD gauging stations throughout Bavaria, Germany.

(*L*, cm), as obtained from each of the GkD gauging stations. Additionally, the daily minimum, maximum and mean air temperatures (*T*, °C) and the daily sunshine durations (*S*, defined by DWD as duration of direct solar radiation at a given location) were derived from the two closest DWD gauging stations for each stream. All data sets carried a time stamp, which we recalculated to obtain the day of year (*D*) as a continuous number. To improve the learning of our ANNs, we employed data normalization, which is a common technique used in machine learning (Han et al., 2011).

We trained and tested all ANNs by inputting data taken from four consecutive days, with the predicted fourth day's daily mean water temperature as the output. This amount of days was found to lead to

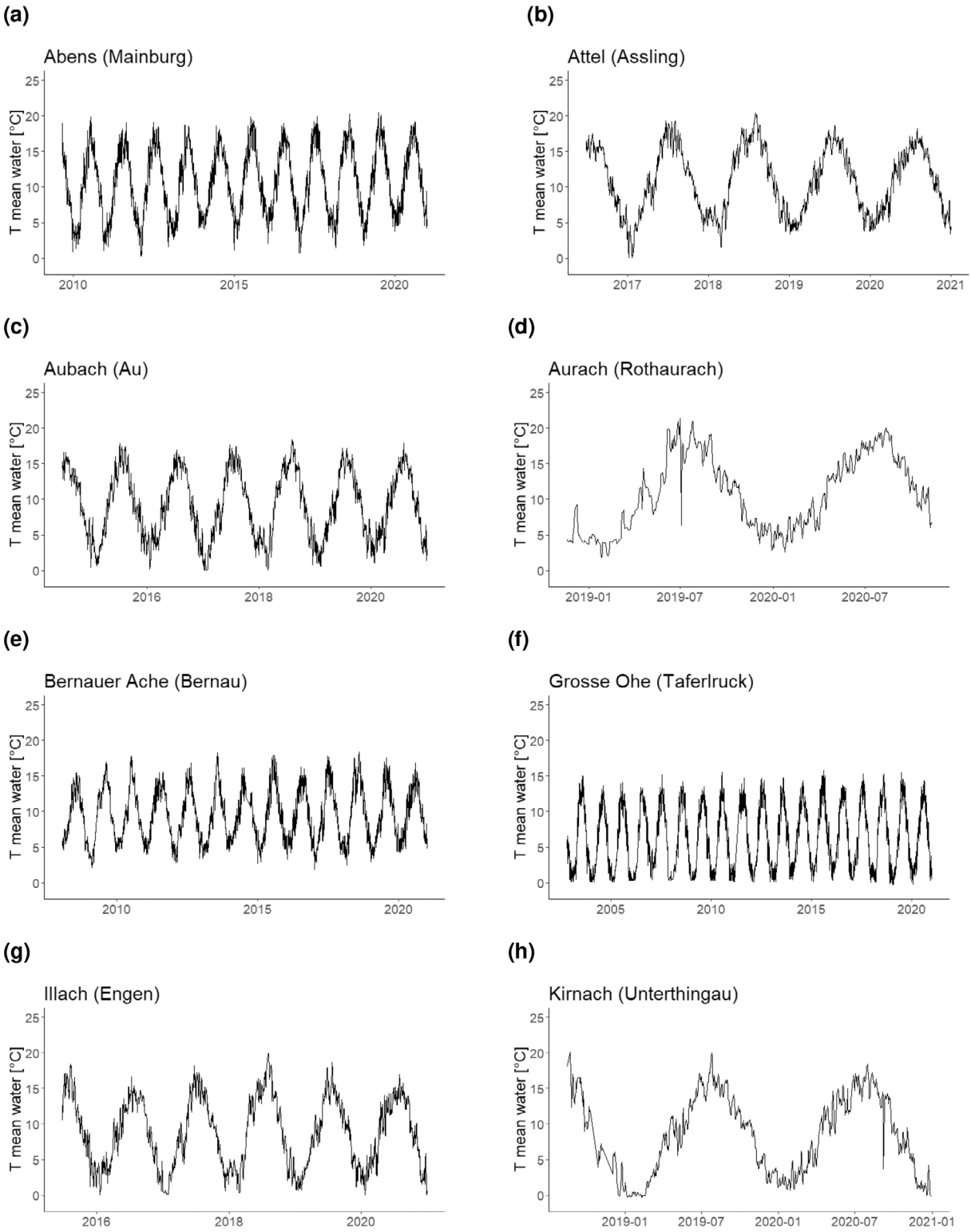

FIGURE 2 Water temperature time series (I). (a) Abens. (b) Attel. (c) Aubach. (d) Aurach. (e) Bernauer Ache. (f) Grosse Ohe. (g) Illach. (h) Kirnach.

a better accuracy compared to ANNs with the input of 1, 2, 3, or 5 consecutive days at a case study in the Bavarian headwater catchment Mähringsbach (Drainas, 2020).

Since not all data were measured continuously, we chose time periods for each stream during which all the input parameters were available (except for Scheine and Soellbach, for which no sunshine data was available). The data used for each stream, the DWD stations used for each GkD gauging station, and the distances between them are presented in Appendix A.4 in Table A2.

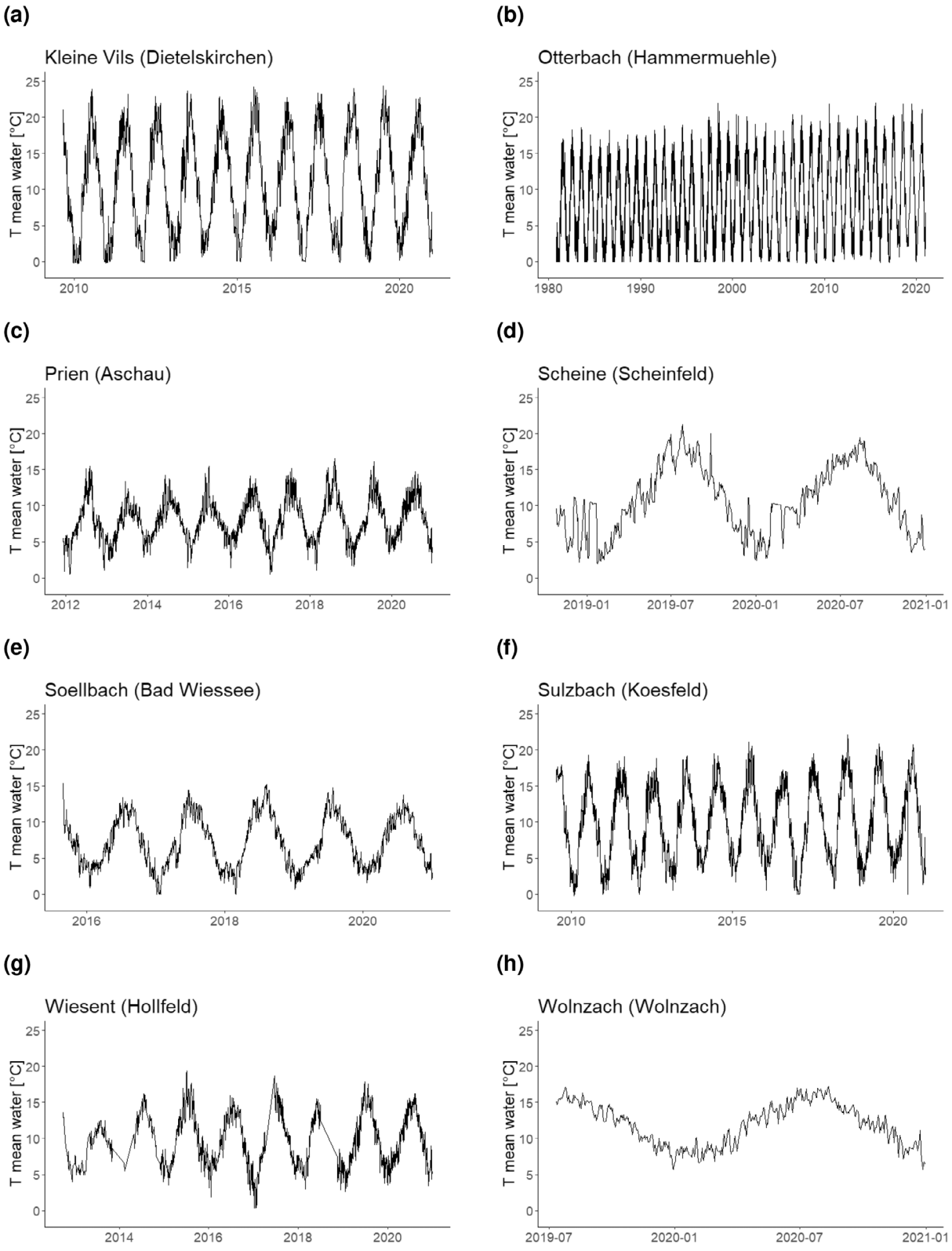

FIGURE 3 Water temperature time series (II). (a) Kleine Vils, (b) Otterbach. (c) Prien. (d) Scheine. (e) Soellbach. (f) Sulzbach. (g) Wiesent. (h) Wolnzach.

To test the suitability of the different input parameters, we trained and tested ANNs with all possible combinations of input parameters for each stream. However, we rejected any combination with no air temperature or date (Mohseni & Stefan, 1999; Zhu, Nyarko, Hadzima-Nyarko, et al., 2019). We used the following combinations (names indicate the input parameters as abbreviated above): *DT*, *DTS*, *DTQ*, *DTL*, *DTQS*, *DTLQ*, *DTLS* and *allinputs* (*DTQLS*).

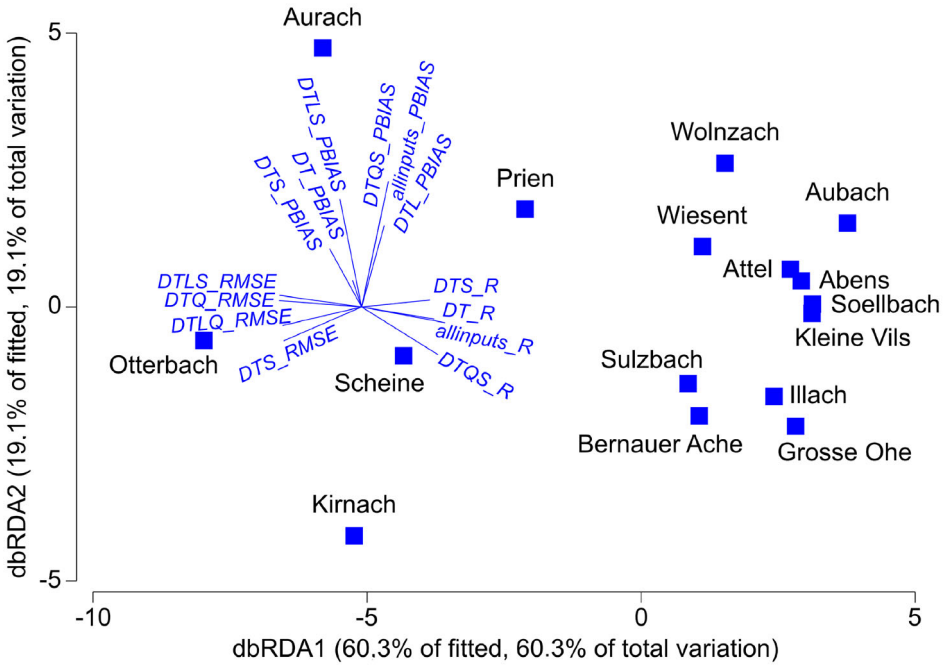

FIGURE 4 distLM-Eval-Eval-plot. Resemblance: D1 Euclidian distance. Strong correlation of RMSE and R along dbRDA axis 1, discrimination of streams by PBIAS on dbRDA axis 2. Input parameters: D, day of the year; T, air temperature; Q, discharge; L, water level; S, sunshine duration; DTQLS, allinputs; R; PBIAS, evaluation metrics as defined in Formulas 1, 2, 3.

2.4 | The modelling approach: Artificial Neural Networks

ANNs are a machine learning approach that is inspired by biological nervous systems (da Silva et al., 2017). They can create output information (in this study water temperature) based on given input information (in this study different input combinations). To do so, they need to learn the relationship between in- and output, for which they need to be trained. Hereby the ANNs have access to the input as well as to the output data and iteratively learn to predict the output based on the input. After the training, the ANNs' ability to predict the output is tested by only presenting the input and comparing the predicted output to the actual output information. In this study, this comparison was conducted by calculating the RMSE.

For the distribution into training and testing phase, the given dataset needs to be divided into a training and a testing dataset. In this study, this was done randomly in a ratio of 90% (training) to 10% (testing) as it is default for an established library (see Appendix A.2 for details).

For this study, we trained and tested nine ANNs for each of the 14 waterbodies, for which all input parameters were available. Additionally, we trained four ANNs for Scheine and four for Soellbach (no sunshine duration available), resulting in a total of 136 ANNs. The ANNs were determined by using search methods from the scikit-learn Python package (Pedregosa et al., 2011) (for a description of the search methods, see Appendix A.1). Optimization was based on the calculated root mean square error (RMSE), which we also employed in this work as an accuracy metric (see Formula 1).

2.5 | Examination of environmental parameters

To determine the influence of the environmental parameters on the prediction accuracy of ANNs, we examined the environmental parameters of two different spatial scales for which an influence can be expected. For that, we used ArcGIS software to create riparian strips upstream of the 16 catchment outlets flanking 5 m left and right (measured from the middle of the stream) of the entire stream length (in the following referred to as "5 m riparian strip"). This width was chosen to capture the direct impact of the landuse on the stream, e.g. in the form of shading. We obtained the stream geometry and catchment spatial maps as input data from the Bavarian Environmental Agency (LfU) and used it to extract various catchment characteristics. Once we had extracted the spatial area of the riparian strips, we intersected the Corine Land Cover (CLC) map (European Union, 2021) with shape files representing the riparian strip and whole catchment area (in the following referred to as "entire catchment resolution") to integrate land use information. The CLC map comprises an inventory of land cover in 44 classes. CLC uses a minimum mapping unit (MMU) of 25 ha for areal phenomena and a minimum width of 100 m for linear phenomena. The MMU for mapping CLC status layers is 25 ha. This means that no objects of less than 25 ha can be present in the database; they are generalized into a neighbouring feature, resulting in >25 ha polygons. Additionally, the minimum mapping width (MMW) of linear elements is 100 m, which means that no objects (e.g., roads, rivers) of less than 100 m width are present in the database.

2.6 | Statistical data analysis

2.6.1 | Environmental dataset

To determine which environmental and land use variables were linked to ANN accuracy and reliability, we assembled an environmental data set which describes the features of the assessed streams on two spatial scales: the entire catchment and the 5 m riparian zone of the stream banks. The data set included the following environmental

parameters for the entire catchment resolution: total length of all tributaries merging into the final river branch (total river length); the length of the longest river branch of the catchment (longest river length); proportion of land use type in the catchment (agriculture, forest, grassland, semi-natural land use, urban and water surfaces); total catchment size; hydrologic parameters calculated over the entire measurement time span, namely mean water level (MW), highest measured water level (HW), lowest measured water level (NW), mean discharge (MQ) and highest measured discharge (HQ); the number of tributaries (tributaries); the number of days for which data was used (DOD); the number of input data points per output data point (IPO); the distance between the GkD station and the closest DWD station (Dist1) and between the GkD station and the second closest DWD station (Dist2).

Regarding the resolution of the 5 m riparian strip, only the land use variables were provided, along with the total riparian strip area.

In Appendix A.4, the description and results of a principal component analysis (PCA) are presented, showing the distribution of the 16 stream sites along the different environmental gradients captured in our environmental dataset.

2.6.2 | Correlation analysis

To determine connections between the prediction accuracy and the environmental parameters, we conducted correlation analysis. For this, we used the calculated accuracy metrics (RMSE, R and PBIAS) for each input combination and each waterbody, once for the entire catchment and once for the 5 m riparian strip resolution. First, we used the Shapiro–Wilk test to examine whether our datasets were normally distributed. We then tested the distribution both for each input combination separately and for each environmental parameter. We then conducted tests to examine the correlation between each input combination and each environmental parameter. If both datasets were normally distributed, we used Pearson product moment correlation. Otherwise, if one or both of the datasets was not normally distributed, we used Spearman rank-order correlation. To visualize the *rho* and *corr* values, we created heatmaps. We conducted all steps of the correlation analysis with RStudio (RStudio Team, 2022; Warnes et al., 2020).

Additionally, we conducted correlation analysis to find connections between reliability and accuracy to assess which accuracy metric is most suitable for displaying the reliability of an ANN's predictions. Details are described in Appendix A.5.

2.6.3 | Distance-based linear model

As a multivariate approach, we used distance based Linear Model (DistLM), which is based on a procedure called "distance based redundancy analysis" (dbRDA) (Legendre & Anderson, 1999) and implements a routine that analyses and models the relationship between a multivariate resemblance matrix and a set of given predictor variables. DistLM is applied as a multivariate multiple regression that models the explanatory significance of the environmental predictor variables via partitioning of variation that facilitates permutation-based significance testing. In our case, we first used the resemblance matrix of the RMSE, R, and PBIAS values of all calculated ANN input combinations and the same data as predictors, to reduce dimensionality. We chose this combination of data, to investigate which of the accuracy metrics and ANN input combinations explained most of the between-stream variability and to identify any redundancy in the three accuracy metrics (Eval-predict). In a subsequent approach, we used the same resemblance matrix but the environmental data set as predictor variables to determine the environmental variables that explain most of the observed variations in the multivariate data set of accuracy metrics of different ANNs (Enviro-predict). For both approaches, we used the DistLM function and redundancy analysis plots (dbRDA-plots) and chose the step-wise method and Adjusted R^2 for model comparison in PRIMER v7 & PERMANOVA+ (Anderson et al., 2008).

3 | RESULTS

3.1 | Measures of model performance

To prioritize the use of the three different accuracy metrics applied in this study, we evaluated which of them is most suitable to also display the reliability of an ANN. Therefore, we conducted correlation analysis between the accuracy and the reliability metrics (for detailed results see Appendix B.2), which resulted in significant correlations between two of the reliability metrics and the RMSE and one significant correlation between reliability metrics and R and PBIAS each.

3.2 | Selection of ANNs and input parameters

The comparison of prediction accuracy of randomly searched ANNs (see RandomizedSearch in Appendix A.1) for all different input combinations (see Tables A3, A4, A5, and A6) showed that the optimal input combination was different for each of the tested streams (see Table 1).

The most frequently used combination of input parameters was DTL (38% of all streams), followed by DTLQ (25% of all streams), allinputs (21%, if sunshine duration was available), DTQS (14%, if sunshine duration was available) and DTQ (6% of all streams) (Table 2 top). The combinations DT, DTS and DTLS were not selected as input combinations with the greatest predictive power in any of the streams. Consequently, the share of individual input parameters in the composition of ANNs was as follows: day of the year and air temperature were identified as input parameters in 100% of all streams, water level was identified in 81% of all streams, discharge was identified in 63% of all streams, and sunshine duration was identified in 36% of the streams for which sunshine duration was available (Table 2 bottom). Using the most accurate input combination for each stream based on the RMSE, the RMSE values ranged between 0.373°C (Aubach) and 1.667°C (Otterbach), R values ranged between 0.997 (Aubach) and 0.958 (Otterbach), and PBIAS values ranged between −0.767% (Kirnach) and 0.112% (Prien). Comparing the input combinations with the

TABLE 1 Evaluation of the most suitable ANN for each waterbody, as determined by RandomizedSearch.

Stream	Inputs	RMSE	*R*	PBIAS
Prien	DTQS	0.733	0.969	0.112
Attel	allinputs	0.453	0.995	0.004
Aubach	DTLQ	0.373	0.997	0.033
Soellbach	DTL	0.419	0.993	−0.090
Bernauer Ache	DTLQ	0.586	0.988	−0.315
Kleine Vils	DTL	0.535	0.997	−0.105
Illach	allinputs	0.503	0.994	−0.136
Otterbach	DTLQ	1.667	0.958	−0.248
Wiesent	DTQS	0.623	0.985	0.158
Sulzbach	DTLQ	0.736	0.990	−0.286
Abens	allinputs	0.468	0.995	−0.084
Aurach	DTL	1.301	0.969	−0.113
Scheine	DTQ	0.920	0.980	−0.676
Grosse Ohe	DTL	0.483	0.994	−0.344
Kirnach	DTL	1.104	0.979	−0.767
Wolnzach	DTL	0.476	0.988	−0.048

Note: Column titles: Stream, name of examined stream; Inputs, input combination used; RMSE, R, PBIAS, evaluation metrics as defined in Formulas 1,2,3. Abbreviations: D, day of the year; DTQLS, allinputs; L, water level; Q, discharge; S, sunshine duration; T, air temperature.

highest accuracy according to the RMSE from each stream, with the combinations of lowest accuracy, the error increased on average by 41% when a random input combination was used, compared to the optimal combination, with a minimum of 5% (Otterbach) and a maximum of 102% (Scheine).

Given the number of ANNs and the accuracy metrics, a DistLM (Eval-predict) was calculated to determine which of the accuracy metrics and ANN input combinations explained the majority of the between-stream variability and to identify redundancy in accuracy metrics. The RMSE and R values were strongly correlated along dbRDA axis 1, implying that the accuracy of calculated ANNs was very similarly reflected by these two metrics (Figure 4). Both allinputs-ANNs of RMSE and R individually explained 58% of the total variability in the dataset according to marginal testing, and the sequential tests furthermore confirmed that the information of R and RMSE of allinputs-ANNs was redundant, as only one of them was included in the best-solution set of variables. However, the PBIAS values calculated on the basis of multiple input combinations were responsible for approximately 35% of the remaining variability in the data set and distinctly discriminated streams on dbRDA axis 2, hence showing that PBIAS provides information that cannot be substituted by the other two used accuracy metrics, and even streams with high accuracy measures, as shown by the small RMSE or high R values, can be subject to under- or overestimation in temperature predictions (Figure 4).

3.3 | ANN accuracy metrics

To determine connections between environmental parameters and ANNs' accuracy, we conducted correlation analyses in the spatial scales of entire catchment as well as 5 m riparian strip resolution.

3.3.1 | Entire catchment

With regard to the entire catchment resolution, we observed the most statistically significant associations between accuracy metrics and environmental parameters for RMSE (34), followed by R (16) and PBIAS (6).

For RMSE (see Figure 5a), we detected significantly positive correlations between all ANN input combinations and total river length as well as between all ANN input combinations and the hydrologic parameters MW and HW. Additionally, four ANN input combinations correlated significantly positively with NW (DT, DTL, DTQ and DTLQ), three ANN input combinations correlated significantly positively with the longest river (DT, DTQ and DTLQ), and three ANN input combinations correlated significantly positively with Dist1 (DTQ, DTQS, DTLS). The R of all input combinations was significantly negatively related to total river length and longest river length (Figure 6a). PBIAS correlated significantly positively with total catchment area (DTLS) and Dist1 (DT) and significantly negatively with DOD (DTLS and DTS) and semi-natural land use (DTS) (Figure 7a).

3.3.2 | 5 m riparian strip

Significant associations between accuracy metrics and environmental parameters in the 5 m riparian strip were most numerous for RMSE (26), followed by R (21) and PBIAS (2). The RMSE values of all ANN input combinations, with the exception of DTLS and DTS, were significantly positively correlated with total riparian strip area (Figure 5b). Also, grassland was significantly positively associated with all ANN input combinations except allinputs and DTQS. Semi-natural land use, in contrast was significantly negatively correlated with RMSE values

TABLE 2 Top: Frequency of input combinations used for the best ANNs as depicted in Table 1. Bottom: Frequency of input parameters used in input combinations in Top.

Input combination	Frequency	Percentage
DTL	6 of 16	38
DTLQ	4 of 16	25
allinputs	3 of 14	21
DTQS	2 of 14	14
DTQ	1 of 16	6
DT	0 of 16	0
DTLS	0 of 14	0
DTS	0 of 14	0
Input Parameter	**Frequency**	**Percentage**
Day of the year (D)	16 of 16	100
Air temperature (T)	16 of 16	100
Water level (L)	13 of 16	81
Discharge (Q)	10 of 16	63
Sunshine duration (S)	5 of 14	36

Note: Sunshine duration was only available for 14 streams.

in DTLS, allinputs and DTQS. Also, the proportion of forest and water surface in the riparian-strip area correlated negatively with RMSE values.

Similarly, R values were significantly negatively related to the total riparian strip area (all ANN combinations except DTS and DTLS), negatively related to grassland (Figure 6b) and positively related to semi-natural land use.

For PBIAS values (see Figure 7b), we only observed a significant negative relationship with semi-natural land use (DT and DTS).

3.4 | Multivariate analysis of environmental predictors of ANN accuracy metrics

To investigate which of the accuracy metrics and ANN input combinations explained most of the between-stream variability and to identify any redundancy in the three accuracy metrics, we conducted a DistLM. We found that the 14 variables depicted in Figure 8 explained a total variation of $R^2 = 0.99601$, Adjusted $R^2 = 0.94014$. 60.5% of the 2-D configuration of the 12 streams was explained by dbRDA axis 1 and 19.05% by dbRDA axis 2.

The DistLM's marginal tests indicated a significant relationship between the multivariate configuration of the streams, based on the three accuracy metrics (RMSE, R, PBIAS) with total river length (prop = 0.35, $p < 0.01$), longest river length (prop = 0.28, $p < 0.01$) and Dist1 (prop = 0.20, $p < 0.05$). Total river length correlated with the negative space of dbRDA axis 1, indicating that decreasing accuracy in terms of RMSE and R (see Figure 8) was significantly correlated with the total length of tributary streams. The streams exemplifying this relationship were the Otterbach in the most negative space of dbRDA axis 1, with a total river length of 41.43 km, contrasting the Aubach with a total river length of 7.91 km, in the upper positive space of dbRDA axis 1. On dbRDA axis 2, streams were mainly separated along a gradient of the parameters: longest river length, Dist1, catchment area as well as proportions of grassland, semi-natural land use and HW. Thus, when relating these findings to the underlying configuration of accuracy metrics, environmental parameters on dbRDA axis 2 were positively associated with overestimation (longest river length, Dist1, catchment area) or underestimation (grassland, semi-natural land use and HW) of water temperature prediction by ANNs.

4 | DISCUSSION

In line with our hypothesis, our results suggest that the accuracy and reliability of ANNs' predictions for single streams are highly dependent on input combination and environmental parameters. To understand how environmental parameters affect ANNs' accuracy and reliability, we analysed a broad range of environmental predictors, showing that river length and water levels, the size of the catchment and open-canopy land use types were particularly negatively associated with ANN accuracy in the streams we tested.

4.1 | Measures of model performance

To prioritize the use of the accuracy metrics RMSE, R, and PBIAS for the evaluation of ANNs, we examined correlations between these metrics and reliability metrics as established in Mohr et al. (2021). We found, that not all accuracy metrics correlated significantly with all reliability metrics, confirming the finding of Mohr et al. (2021), that the use of accuracy metrics alone is insufficient and should be supplemented by reliability metrics.

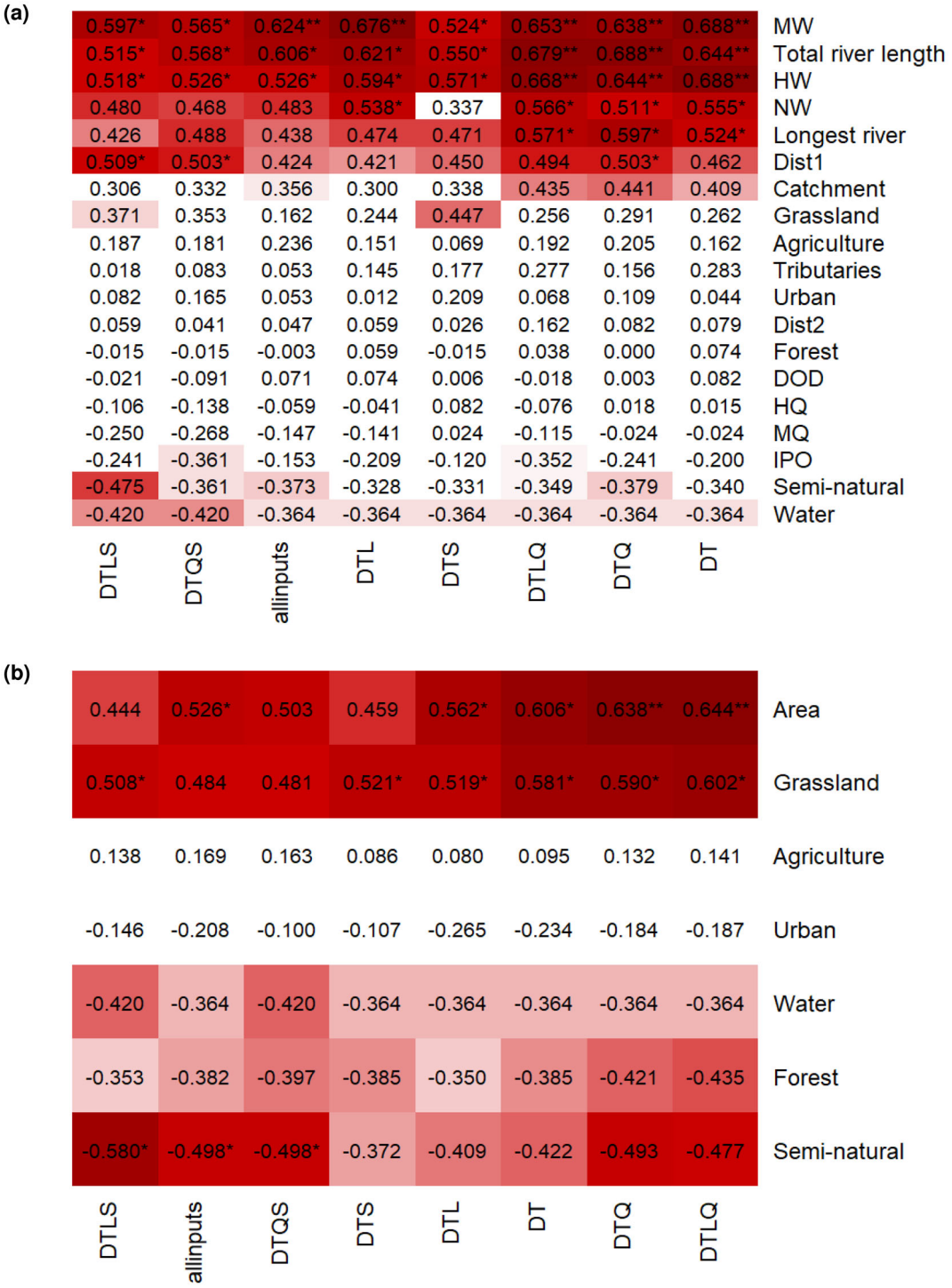

FIGURE 5 (a) RMSE entire catchment; (b) RMSE 5 m riparian strip. Increasing intensity of red colour indicates increasing correlation (positive as well as negative). Significance is marked with $*p < 0.05$ and $**p < 0.01$. Input parameters (*x*-axis): D, day of the year; T, air temperature; Q, discharge; L, water level; S, sunshine duration; DTQLS, allinputs. Environmental parameters (*y*-axis): as described in Section 2.6.

Still, we can conclude that as accuracy metric, the RMSE was the most suitable one to reflect the reliability of an ANN, due to two significant correlations with reliability metrics as opposed to one significant correlation for R and PBIAS each. We also observed that the RMSE had a greater resolution and hence contributed more significant relationships with environmental parameters than R, probably because it had a higher potential to reflect the high-resolution dynamics of hydrologic parameters. This further confirmed the plausibility of its

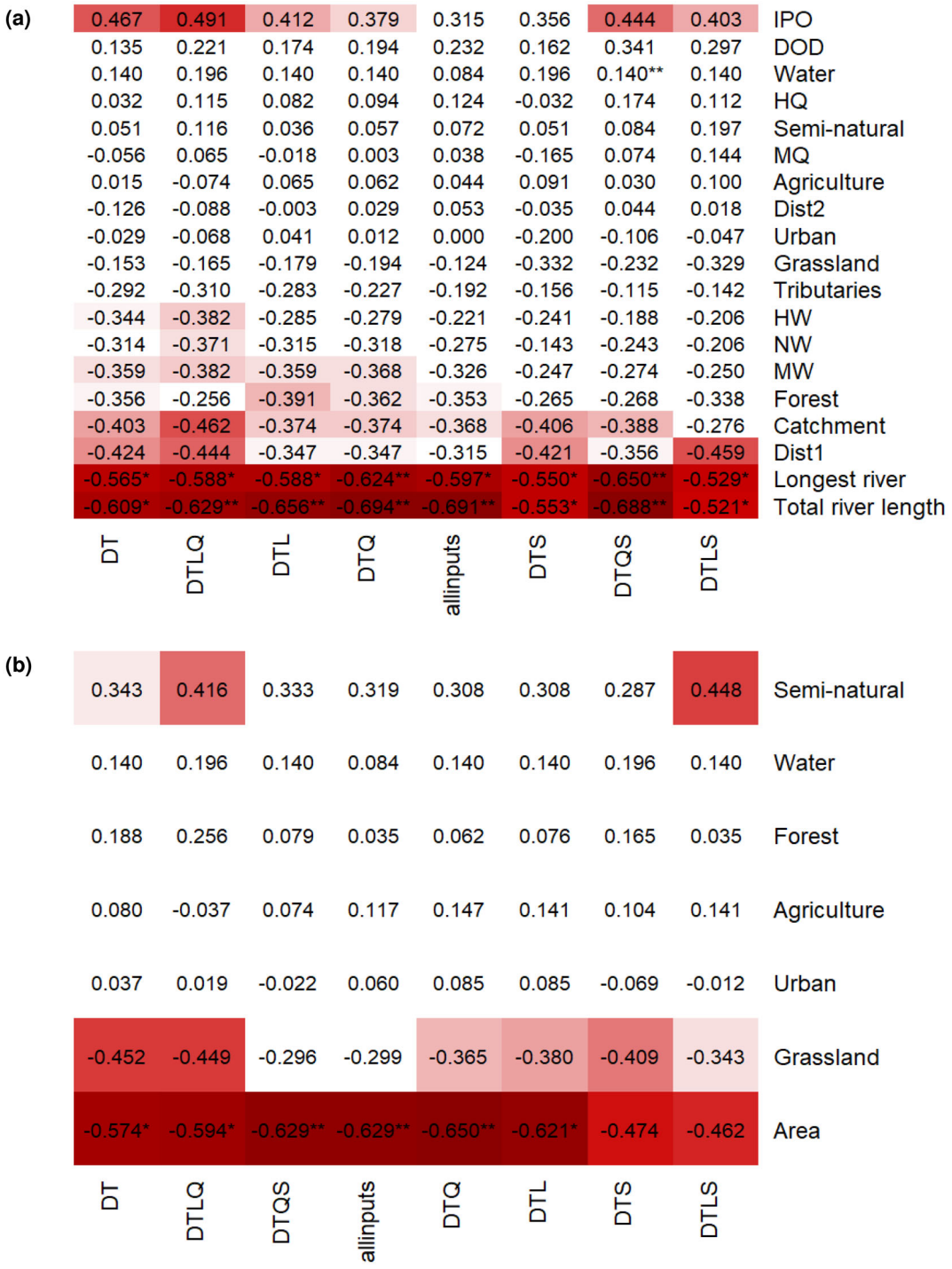

FIGURE 6 (a) R entire catchment; (b) R 5 m riparian strip. Increasing intensity of red colour indicates increasing correlation (positive as well as negative). Significance is marked with $^{*}p < 0.05$ and $^{**}p < 0.01$. Input parameters (*x*-axis): D, day of the year; DTQLS, allinputs; L, water level; Q, discharge; S, sunshine duration; T, air temperature; Environmental parameters (*y*-axis): as described in Section 2.6.

frequent use for measuring the accuracy of water temperature prediction with ANNs (Ahmadi-Nedushan et al., 2007; Caissie et al., 1998; Chenard & Caissie, 2008; Cho & Lee, 2012; Feigl et al., 2021; Graf et al., 2019; Hadzima-Nyarko et al., 2014; Qiu et al., 2020; Quan et al., 2020; Rabi et al., 2015; Rahmani et al., 2020; Rehana, 2019; St-Hilaire et al., 2000; Zhu, Nyarko, Hadzima-Nyarko, et al., 2019).

We additionally employed PBIAS as an accuracy metric, which is unusual for water temperature prediction with ANNs. Although we saw advantages of including the PBIAS due to the different aspects of model performance it highlights, in this study we were not able to find any general significant correlations between the assessed environmental parameters and the PBIAS. This might be because the PBIAS

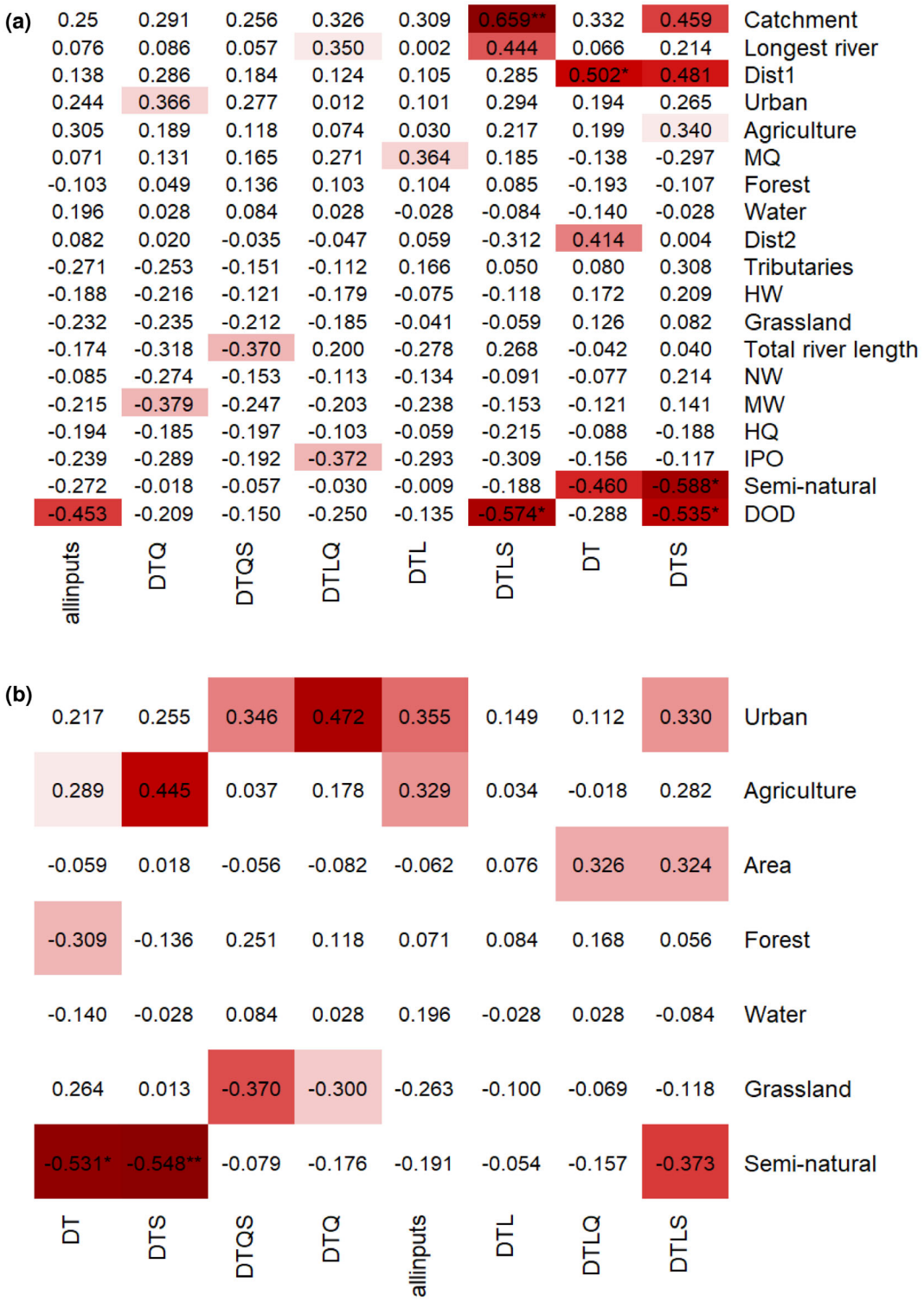

FIGURE 7 (a) PBIAS entire catchment; (b) PBIAS 5 m riparian strip. Increasing intensity of red colour indicates increasing correlation (positive as well as negative). Significance is marked with $*p < 0.05$ and $**p < 0.01$. Input parameters (*x*-axis): D, day of the year; DTQLS, allinputs; L, water level; Q, discharge; S, sunshine duration; T, air temperature; Environmental parameters (*y*-axis): As described in Section 2.6.

in general reflects variation in two directions, but in our study the direction of estimation (over- or underestimation) did not necessarily correlate with the examined environmental parameters in only one direction. The overestimation was pronounced for Kirnach, a stream with a very high proportion of grassland (72.58%) and a very low proportion of semi-natural land cover (0.01%). In contrast, underestimation of water temperature was pronounced for Aurach, a long stream with a large catchment. These findings were also confirmed by the

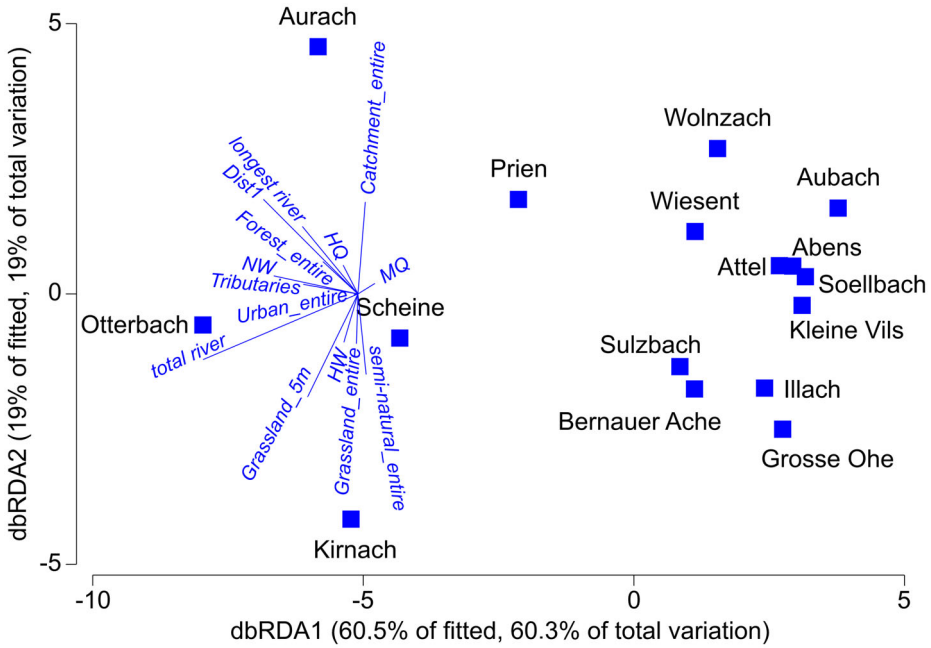

FIGURE 8 distLM-Eval-Enviro-plot, Resemblance: D1 Euclidian distance, Correlation between total river length and negative space of dbRDA axis 1, discrimination of streams by longest river length, Dist1, catchment area, proportions of grassland and semi-natural land use, and HW along dbRDA axis 2. Environmental parameters: Total river length: Sum of lengths of all contributing rivers; Longest river length: Length of the longest contributing river; Land use: agriculture, forest, grassland, semi-natural, urban, water; Catchment: Catchment size of all contributing catchments; Area: Total buffer area; MW: Mean water level; HW: Highest measured water level; NW: Lowest measured water level; MQ: Mean discharge; HQ: Highest measured discharge; Tributaries: Number of tributaries; DOD: Number of days for which data was used; IPO: Number of input data points per output data point; Dist1: Distance between GkD station and DWD station 1; Dist2: Distance between GkD station and DWD station 2. Resolution: entire: Entire catchment; 5 m 5 m riparian strip.

DistLM analysis of environmental predictors of evaluation metrics, in which PBIAS/overestimation was associated with high proportions of grassland, particularly in the 5 m riparian strip. Consequently, it would be advisable to carefully check for both over- and underestimation of water temperature prediction, particularly in catchments with high proportions of open-canopy landscape.

4.2 | Input parameters

The most striking finding of this study was that the input combination with the highest accuracy was a stream-specific set of input parameters, suggesting that the optimal input combination cannot be generalized across streams. As important asset, this study used a systematic procedure of training and testing ANNs with different sets of input parameters, which provided us with the opportunity to compare ANN accuracies within single streams. While other studies like Feigl et al. (2021) previously identified that the input combination has an effect on ANN performance in general, our finding, that the optimal input combination is stream-specific, adds an important new insight to this field, which can help to make stream water temperature predictions more accurate in the future. As we were able to show, the error in the prediction (RMSE) could increase to > 100% in a single stream if a random input combination was used instead of the optimal input combination. Even when using the allinputs combination, the error increased by up to 34%, indicating that allinputs might be more accurate than a random input combination, but still not as accurate as if the combination was determined systematically. This result is in line with the "explosion" of Myth #7 in Maier et al. (2023), where it is stated that an increase in the number of input variables does not necessarily improve model performance, but that these variables need to be selected carefully. Clearly, the search for the optimal input combination is time consuming compared to a fixed procedure using a set of pre-defined input variables. Hence, for supporting the application of ANNs based on our results, we provide a flow chart to facilitate decision-making along the process of water temperature prediction with ANNs (see Figure 9).

Comparing the RMSE values from Table 1 to previous studies that predicted water temperatures with ANNs (sota-range 0.46°C to 1.58°C), only one stream (Otterbach, RMSE = 1.667°C) had an RMSE slightly higher than the sota-range. Further, 12 streams had an RMSE within the sota-range and three streams had RMSE values even lower than the sota-range, namely Attel (RMSE = 0.453°C), Aubach (RMSE = 0.373°C), and Soellbach (RMSE = 0.419°C). To the best of our knowledge, the RMSE values of Attel, Aubach and Soellbach were the smallest ever reported for stream water temperature prediction using ANNs.

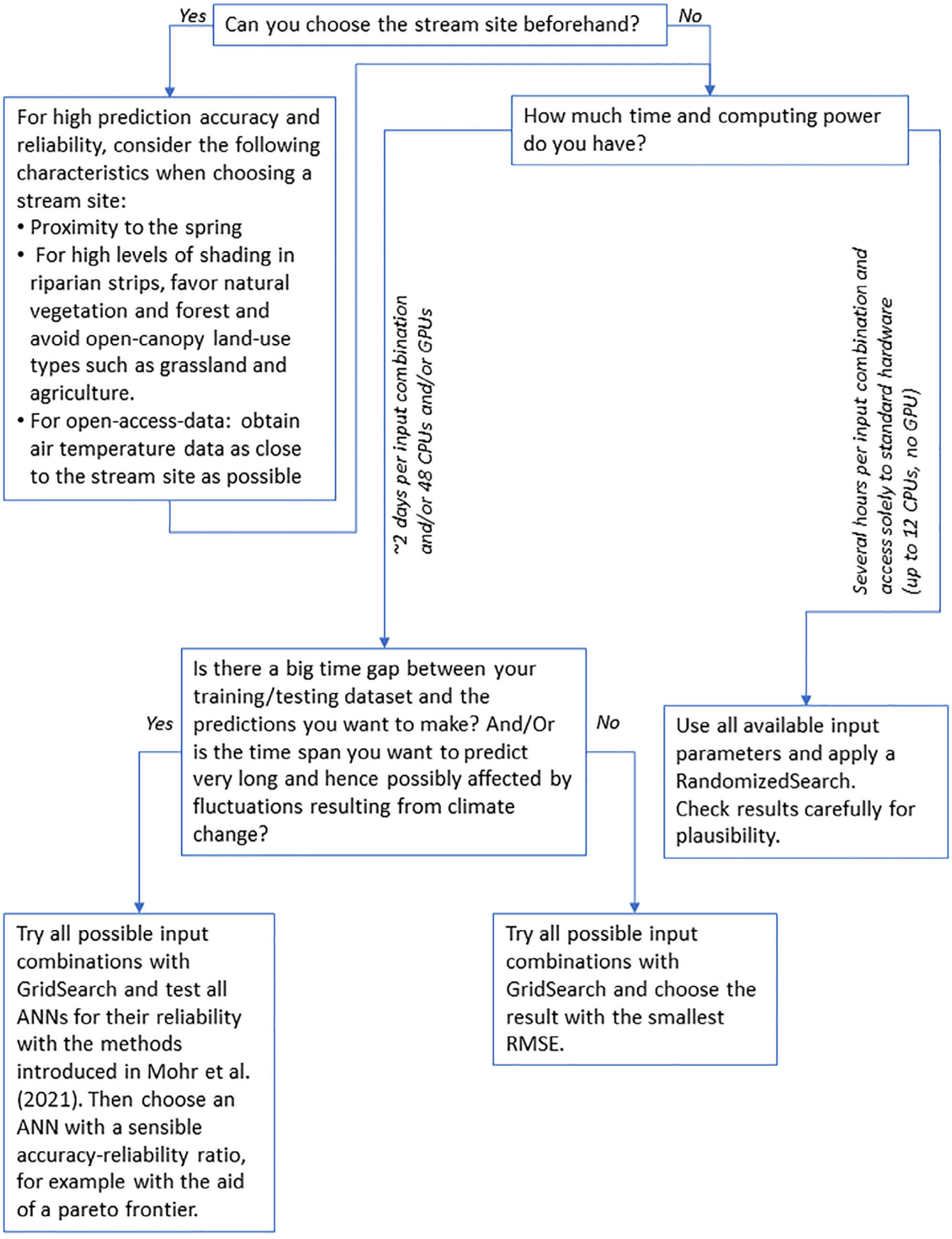

FIGURE 9 Flow chart with recommendations on stream-specific artificial neural network-development.

Based on Zhu, Nyarko, Hadzima-Nyarko, et al. (2019), we expected a minor role of discharge in explaining temperature, since they state that discharge plays a minor role in stream water temperature prediction compared to the day of the year and that discharge's importance increases for high-altitude catchments. Still, in all of the streams of our study, the most accurate ANNs all had water level and/or discharge as inputs. Unfortunately, Zhu, Nyarko, Hadzima-Nyarko, et al. (2019) did not consider water level, which hinders direct comparison with our results. Nevertheless, based on our results, we suggest using at least one hydrologic input parameter for water temperature prediction with ANNs, while we cannot generalize the recommendation to a specific hydrologic parameter, since this is highly stream specific. Still, we conclude that no unique optimal input combination exists for each stream.

4.3 | Environmental influences on water temperature prediction

Given the high specificity of input combinations we determined for individual streams, it was a key interest of this study to identify stream environmental conditions that govern the accuracy of ANNs. In light of climate change, such knowledge is also highly relevant for deducing mitigation and management strategies in streams related to securing high oxygen concentrations (Piatka et al., 2021), endangered fish populations (Wild et al., 2023), and temperature refuges (Kuhn et al., 2021; Mejia et al., 2023). In contrast to existing approaches, which mainly consider hyperparameter tuning and dataset length, the associations between environmental parameters and ANN accuracy allow a more mechanistic and realistic assessment of model applicability at individual stream sites, as demonstrated for our datasets. Several significant correlations between environmental parameters and prediction accuracy of ANNs were identified, suggesting key influences of catchment hydrological variables.

Specifically, the accuracy of ANNs (RMSE and R) was strongly related to total river and longest river length, total catchment area, and the hydrological parameters MW, HW, and NW, indicating a decrease in ANN accuracy with increasing river length, catchment size, and water level.

Since stream water temperatures are defined by complex and dynamic physico-chemical, hydrologic and atmospheric processes and not solely based on air temperature (Caissie, 2006; Leach et al., 2023), a possible explanation for the strong negative relationship between ANN accuracy and river length and catchment area could be the increase of air-temperature-unrelated complex influences along the flow path of streams. Beginning at the spring, the water has a specific temperature, depending on its origin and the distance to its spring. As the stream water passes through the landscape, energy exchange is influenced by advective fluxes like evaporation or longitudinal changes in advection and radiation due to changes in vegetation (Coats & Jackson, 2020; Leach et al., 2023). Energy is added by river bank and bed friction, and contact with the atmosphere increases, as do the radiative fluxes (Dan Moore et al., 2005; Kuhn et al., 2021; Webb et al., 2008). Hence, with increasing river length, the potential number of complex influences increases and thus, the accuracy of the water temperature predictions decreases. This is especially pronounced for models like ANNs, which do not receive additional information on catchment-size related variables but have to learn in the context of local input parameters, measured at the gauging station.

As with river length and catchment size, higher levels of HW, NW and MW were associated with a lower prediction accuracy (RMSE) of ANNs. The relationship between extreme water levels (HW) and ANN accuracy is due to difficulties in predicting the temperatures of water sources entering the stream along its flowpath (e.g. groundwater, hyporheic water, precipitation, anthropogenic water influxes (Nelson & Palmer, 2007; Webb et al., 2008). During spates and high-water events, these water sources contribute different relative quantities to total water volume, and temperature mixing during high water events is then presumably more difficult for ANNs to predict. Additionally, it has been shown that air-water temperature relationships are stronger and more sensitive for flows below median levels (Webb et al., 2003), likely because high water levels lead to a lower water-atmosphere interaction of the surface area compared to total water volume, influencing radiation influx and sensible heat transfer. As a result, depending on surrounding atmospheric temperatures, energy fluxes are often easier to predict for smaller water volumes, which explains the higher prediction accuracy for lower MW and NW values of streams. Hence, the connection between increasing water levels, in particular the HW values and decreasing accuracy in water temperature prediction by ANNs, seems plausible and should be considered when predicting water temperatures in streams during periods of high water.

We found that the land use types semi-natural, forest and water bodies had a positive effect on ANN accuracy. Further, our results showed that high proportions of grassland in the 5 m riparian strip (but not on the entire catchment resolution) correlated with decreasing accuracy (RMSE) in water temperature prediction.

The land use surrounding a stream has a strong influence on its temperature regime and humidity, which controls the water-atmosphere interaction (Webb & Zhang, 1997). It can be assumed that high proportions of grassland facilitated heat-induced evaporation, which can lead to cooling effects especially during high temperature phases (Ouellet & Caissie, 2023), inducing a paradoxical relationship between air temperature (increasing) and water temperature (decreasing). In low temperature phases, this effect is not induced, resulting in an inconsistent relationship between air and water temperature, hence potentially reducing the accuracy of water temperature predictions based on air temperature data.

In general, open-canopy land use such as grassland involves higher levels of radiation and heat fluxes due to a lack of shading and temperature buffering through a micro-climate of complex riparian vegetation. As solar radiation is the most important component of heat transfer in streams (Webb & Zhang, 1997, 1999), accounting for

70% of non-advective heat fluxes in a stream (Webb et al., 2008), open-canopy land use forms are associated with higher air temperatures and lower humidity, which can in turn result in more pronounced temperature extremes and drought conditions in streams. For example, Rutherford et al. (2004) and Ebersole et al. (2003) attributed a maximum temperature decrease of 4°C downstream of shaded areas to the effect of riparian vegetation. In a simulation study, Wondzell et al. (2019) determined that shading through a mature forest can account for a decrease of water temperature of 8°C. Johnson (2004) quantified the net energy transfer in July in a stream in Oregon. Non-shaded, the water temperature gained 580 W/m^2, but fully shaded, the stream's water lost 149 W/m^2. Hoess et al. (2022) found that shading by coniferous vegetation could even compensate for a temperature increase caused by pond effluents. Also, without shading, conduction between water and heated alluvial substrates is an often underestimated process influencing stream water temperatures, particularly under forest harvest scenarios (Brown, 1969; Johnson & Jones, 2000). Hence, riparian shading appears to be of paramount importance for controlling and regulating stream water temperatures. Our findings further demonstrated that for the prediction of water temperatures using an air-water-temperature relationship, the land use in the proximate riparian surroundings (in our case the 5 m riparian strip) seemed more important than the catchment's global land use. Also, Kail et al. (2021) found that large trees in the 10 m riparian strip are a better predictor of water temperature than the width of riparian strips (in their case 30 m), due to the presence of large trees that provided direct shade for the streams and hence cooled the stream water highly effectively. As we showed that prediction accuracy (RMSE) was higher in streams with higher proportions of forest and semi-natural land use (5 m riparian strip) and semi-natural land use and water body area (entire catchment), it can be assumed that riparian shade stabilizes water temperatures, hence facilitating more accurate prediction, as water temperatures are likely more linearly and consistently related to atmospheric temperatures. Also the proportion of water bodies is likely related to prediction accuracy, due to their temperature-buffering properties in the catchment. Our results imply that larger proportions of open-canopy land use forms and the associated higher radiation and low levels of shading can lead to high levels of temperature variability, potentially hampering ANN accuracy and reliability. Consequently, we advise greater caution when using ANNs for streams in open-canopy landscapes.

5 | CONCLUSIONS

We conclude the following for water temperature prediction in streams with ANNs, based on open-access data:

1. It is possible to use open-access data for water temperature predictions within the sota-range.
   a. The use of open-access data, however, comes with the problem that there is only a limited number of parameters. Hence, the choice of streams for which the water temperature is to be predicted is crucial for the accuracy and reliability of the predictions.
   b. For an optimal outcome, all available input parameters should be tested for their suitability (see recommendations in Figure 9).
2. If water temperature is to be predicted for a specific stream, it might not be sufficient to use open-access data, especially if the stream is characterized by specific environmental parameters, which reduce the accuracy and reliability of water temperature prediction.
3. If the ANN is intended to predict water temperature for a future or past time with different climatic conditions compared to present ones, not only the accuracy but also the reliability of the ANN should be considered in the choice of architecture and input parameters (see recommendations in Figure 9). If it is not possible to test reliability, the RMSE is a good (but not in itself sufficient) predictor of ANN reliability and should hence be used.

Our findings highlight that water temperature predictions are more accurate and reliable in headwater streams closer to their source, especially if adjacent land use comprises forests and natural riparian vegetation that lack anthropogenic influences. The finding that ANN prediction accuracy is distinctly compromised by disturbances in the riparian cover, which commonly accumulate along a river's course, leads us to conclude that the lower ANN accuracy reflects the increasing disturbances in the air-water-temperature relationship. We propose that measures of ANN accuracy, as a proxy for an inconsistent air-water-temperature relationship, could even be used to indicate a functional and resilient water temperature regime in headwater streams. Given the importance of small headwater streams and spring ecosystems as refuges and highly specialized environments that feature a broad width of unique and sensitive species requiring special protection (Cantonati et al., 2012; Richardson, 2019), ANN accuracy measures could serve as an indicative tool to identify, evaluate and monitor headwater streams with regard to their temperature integrity and to support decision making regarding where and how to best protect these unique environments. Further, this research highlights that anthropogenic and, specifically, land-use-derived disturbances along stream ecosystems affect stream water temperatures and will consequently exacerbate the climate-change-associated warming of stream water. We have therefore added highly relevant information to the use of ANNs to predict stream water temperatures. In combination with climate change projections, ANNs could prove to be a cost-efficient and invaluable resource for decision makers to use when assessing future developments in stream water temperatures, aiding the evaluation and prioritization of restoration, renaturation and adaptation measures in streams.

ACKNOWLEDGEMENTS

This work was supported by the Bavarian State Ministry of Science and the Arts in the AquaKlif project within the Bavarian Climate Research Network (bayklif) and by the DFG Research Training Group

on Continuous Verification of Cyber-Physical Systems (GRK 2428). We thank Jan Křetínský and Maximilian Weininger from the Chair for Foundations of Software Reliability and Theoretical Computer Science (Technical University of Munich) for the valuable discussions on model assessment as well as the Proofreading Service of the TUM Graduate School for a final cross-check of the manuscript. Open Access funding enabled and organized by Projekt DEAL.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Konstantina Drainas https://orcid.org/0000-0001-6771-2123
Lisa Kaule https://orcid.org/0000-0002-5881-7561
Stefanie Mohr https://orcid.org/0000-0002-8630-3218
Bhumika Uniyal https://orcid.org/0000-0002-3841-8184
Romy Wild https://orcid.org/0000-0002-4814-6215
Juergen Geist https://orcid.org/0000-0001-7698-3443

REFERENCES

Ahmadi-Nedushan, B., St-Hilaire, A., Ouarda, T. B. M. J., Bilodeau, L., Robichaud, É., Thiémonge, N., & Bobée, B. (2007). Predicting river water temperatures using stochastic models: Case study of the Moisie River (Québec, Canada). *Hydrological Processes*, *21*(1), 21–34.

Anderson, M. J., Gorley, R., & Clarke, K. (2008). *PERMANOVA+ for PRIMER: Guide to software and statistical methods*. PRIMER-E.

Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K.-R., & Samek, W. (2015). On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLoS One*, *10*(7), e0130140.

Baehrens, D., Schroeter, T., Harmeling, S., Kawanabe, M., Hansen, K., & Müller, K.-R. (2010). How to explain individual classification decisions. *The Journal of Machine Learning Research*, *11*, 1803–1831.

Benyahya, L., Caissie, D., St-Hilaire, A., Ouarda, T. B., & Bobée, B. (2007). A review of statistical water temperature models. *Canadian Water Resources Journal*, *32*(3), 179–192.

Brown, G. W. (1969). Predicting temperatures of small streams. *Water Resources Research*, *5*(1), 68–75.

Caissie, D. (2006). The thermal regime of rivers: A review. *Freshwater Biology*, *51*(8), 1389–1406.

Caissie, D., El-Jabi, N., & St-Hilaire, A. (1998). Stochastic modelling of water temperatures in a small stream using air to water relations. *Canadian Journal of Civil Engineering*, *25*(2), 250–260.

Caldwell, P., Segura, C., Gull Laird, S., Sun, G., McNulty, S. G., Sandercock, M., Boggs, J., & Vose, J. M. (2015). Short-term stream water temperature observations permit rapid assessment of potential climate change impacts. *Hydrological Processes*, *29*(9), 2196–2211.

Cantonati, M., Füreder, L., Gerecke, R., Jüttner, I., & Cox, E. J. (2012). Crenic habitats, hotspots for freshwater biodiversity conservation: Toward an understanding of their ecology. *Freshwater Science*, *31*(2), 463–480.

Chen, Y. D., Carsel, R. F., McCutcheon, S. C., & Nutter, W. L. (1998). Stream temperature simulation of forested riparian areas: I. Watershed-scale model development. *Journal of Environmental Engineering*, *124*(4), 304–315.

Chenard, J.-F., & Caissie, D. (2008). Stream temperature modelling using artificial neural networks: Application on catamaran brook, New Brunswick, Canada. *Hydrological Processes*, *22*(17), 3361–3372.

Cho, H.-Y., & Lee, K.-H. (2012). Development of an air–water temperature relationship model to predict Climate-induced future water temperature in estuaries. *Journal of Environmental Engineering*, *138*(5), 570–577.

Coats, W. A., & Jackson, C. R. (2020). Riparian canopy openings on mountain streams: Landscape controls upon temperature increases within openings and cooling downstream. *Hydrological Processes*, *34*(8), 1966–1980.

Crisp, D. T., & Howson, G. (1982). Effect of air temperature upon mean water temperature in streams in the north Pennines and English Lake District. *Freshwater Biology*, *12*(4), 359–367.

da Silva, I. N., Hernane Spatti, D., Andrade Flauzino, R., Liboni, L. H. B., & dos Reis Alves, S. F. (2017). *Artificial neural networks*. Springer International Publishing.

Dan Moore, R., Spittlehouse, D., & Story, A. (2005). Riparian microclimate and stream temperature response to forest harvesting: A review 1. *JAWRA Journal of the American Water Resources Association*, *41*(4), 813–834.

Drainas, K. (2020). Prediction of stream water and hyporheic temperature in the context of local climate change: A case study at the bavarian mähringsbach, fichtel mountains. [Unpublisahed Master's thesis], Technical University of Munich.

Ebersole, J. L., Liss, W. J., & Frissell, C. A. (2003). Cold water patches in warm streams: Physicochemical characteristics and the influence of shading 1. *JAWRA Journal of the American Water Resources Association*, *39*(2), 355–368.

Erhan, D., Bengio, Y., Courville, A., & Vincent, P. (2009). Visualizing higher-layer features of a deep network. *University of Montreal*, *1341*(3), 1.

European Union. (2021). *Copernicus Land Monitoring Service 2021*. European Environment Agency (EEA).

Feigl, M., Lebiedzinski, K., Herrnegger, M., & Schulz, K. (2021). Machine-learning methods for stream water temperature prediction. *Hydrology and Earth System Sciences*, *25*(5), 2951–2977.

Graf, R., Zhu, S., & Sivakumar, B. (2019). Forecasting river water temperature time series using a wavelet–neural network hybrid modelling approach. *Journal of Hydrology*, *578*, 124115.

Gupta, H. V., Sorooshian, S., & Yapo, P. O. (1999). Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration. *Journal of Hydrologic Engineering*, *4*(2), 135–143.

Hadzima-Nyarko, M., Rabi, A., & Šperac, M. (2014). Implementation of artificial neural networks in modeling the water-air temperature relationship of the river Drava. *Water Resources Management*, *28*(5), 1379–1394.

Han, J., Pei, J., & Kamber, M. (2011). *Data mining: Concepts and techniques*. Elsevier.

Harvey, R., Lye, L., Khan, A., & Paterson, R. (2011). The influence of air temperature on water temperature and the concentration of dissolved oxygen in Newfoundland Rivers. *Canadian Water Resources Journal*, *36*(2), 171–192.

Hoess, R., Generali, K. A., Kuhn, J., & Geist, J. (2022). Impact of fish ponds on stream hydrology and temperature regime in the context of freshwater pearl mussel conservation. *Watermark*, *14*(16), 2490.

Huang, X., Kroening, D., Ruan, W., Sharp, J., Sun, Y., Thamo, E., Wu, M., & Yi, X. (2020). A survey of safety and trustworthiness of deep neural networks: Verification, testing, adversarial attack and defence, and interpretability? *Computer Science Review*, *37*, 100270.

IPCC. (2022). Climate change 2022: Impacts, adaptation and vulnerability. Contribution of working group II to the sixth assessment report of the intergovernmental panel on Climate change. In H.-O. Pörtner, D. C. Roberts, M. Tignor, E. S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, & B. Rama (Eds.). Cambridge University Press.

Johnson, S. L. (2004). Factors influencing stream temperatures in small streams: Substrate effects and a shading experiment. *Canadian Journal of Fisheries and Aquatic Sciences*, *61*(6), 913–923.

Johnson, S. L., & Jones, J. A. (2000). Stream temperature responses to forest harvest and debris flows in western cascades, Oregon. *Canadian Journal of Fisheries and Aquatic Sciences*, *57*(S2), 30–39.

Kail, J., Palt, M., Lorenz, A., & Hering, D. (2021). Woody buffer effects on water temperature: The role of spatial configuration and daily temperature fluctuations. *Hydrological Processes*, *35*(1), e14008.

Kothandaraman, V. and Evans, R. L. (1972). *Use of air-water relationships for predicting water temperature*. Illinois State Water Survey.

Krider, L. A., Magner, J. A., Perry, J., Vondracek, B., & Ferrington, L. C. (2013). Air-water temperature relationships in the trout streams of southeastern Minnesota's carbonate-sandstone landscape. *JAWRA Journal of the American Water Resources Association*, *49*(4), 896–907.

Kuhn, J., Casas-Mulet, R., Pander, J., & Geist, J. (2021). Assessing stream thermal heterogeneity and cold-water patches from UAV-based imagery: A matter of classification methods and metrics. *Remote Sensing*, *13*(7), 1379.

Leach, J. A., Kelleher, C., Kurylyk, B. L., Moore, R. D., & Neilson, B. T. (2023). A primer on stream temperature processes. *Wiley Interdisciplinary Reviews Water*, 10, e1643.

Legendre, P., & Anderson, M. J. (1999). Distance-based redundancy analysis: Testing multispecies responses in multifactorial ecological experiments. *Ecological Monographs*, *69*(1), 1–24.

Maier, H. R., Galelli, S., Razavi, S., Castelletti, A., Rizzoli, A., Athanasiadis, I. N., Sànchez-Marrè, M., Acutis, M., Wu, W., & Humphrey, G. B. (2023). Exploding the myths: An introduction to artificial neural networks for prediction and forecasting. *Environmental Modelling & Software*, *167*, 105776.

Mejia, F. H., Ouellet, V., Briggs, M. A., Carlson, S. M., Casas-Mulet, R., Chapman, M., Collins, M. J., Dugdale, S. J., Ebersole, J. L., Frechette, D. M., Fullerton, A. H., Gillis, C. A., Johnson, Z. C., Kelleher, C., Kurylyk, B. L., Lave, R., Letcher, B. H., Myrvold, K. M., Nadeau, T. L., … Torgersen, C. E. (2023). Closing the gap between science and management of cold-water refuges in rivers and streams. *Global Change Biology*, *29*, 5482–5508.

Mohr, S., Drainas, K., & Geist, J. (2021). Assessment of neural networks for stream-water-temperature prediction. In *20th IEEE international conference on machine learning and applications (ICMLA)*. IEEE.

Mohseni, O., & Stefan, H. G. (1999). Stream temperature/air temperature relationship: A physical interpretation. *Journal of Hydrology*, *218*(3–4), 128–141. PII: S0022169499000347.

Moriasi, D. N., Arnold, J. G., van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, *50*(3), 885–900.

Nelson, K. C., & Palmer, M. A. (2007). Stream temperature surges under urbanization and Climate change: Data, models, and responses. *JAWRA Journal of the American Water Resources Association*, *43*(2), 440–452.

Ouellet, V., & Caissie, D. (2023). Towards a better understanding of the evaporative cooling of rivers: Case study for the little Southwest Miramichi river (New Brunswick, Canada). *Canadian Water Resources Journal/Revue Canadienne Des Ressources Hydriques*, *48*, 1–17.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Piatka, D. R., Wild, R., Hartmann, J., Kaule, R., Kaule, L., Gilfedder, B., Peiffer, S., Geist, J., Beierkuhnlein, C., & Barth, J. A. (2021). Transfer and transformations of oxygen in rivers as catchment reflectors of continental landscapes: A review. *Earth-Science Reviews*, *220*, 103729.

Pilgrim, J. M., Fang, X., & Stefan, H. G. (1998). Stream temperature CORRELATIONS with air temperatures IN Minnesota: Implications for CLIMATE warming. *JAWRA Journal of the American Water Resources Association*, *34*(5), 1109–1121.

Piotrowski, A. P., Napiorkowski, M. J., Napiorkowski, J. J., & Osuch, M. (2015). Comparing various artificial neural network types for water temperature prediction in rivers. *Journal of Hydrology*, *529*, 302–315.

Qiu, R., Wang, Y., Wang, D., Qiu, W., Wu, J., & Tao, Y. (2020). Water temperature forecasting based on modified artificial neural network methods: Two cases of the Yangtze River. *The Science of the Total Environment*, *737*, 139729.

Quan, Q., Hao, Z., Xifeng, H., & Jingchun, L. (2020). Research on water temperature prediction based on improved support vector regression. *Neural Computing and Applications*, *34*, 1–10.

Rabi, A., Hadzima-Nyarko, M., & Šperac, M. (2015). Modelling river temperature from air temperature: Case of the river Drava (Croatia). *Hydrological Sciences Journal*, *60*(9), 1490–1507.

Rahmani, F., Lawson, K., Ouyang, W., Appling, A., Oliver, S., & Shen, C. (2020). Exploring the exceptional performance of a deep learning stream temperature model and the value of streamflow data. *Environmental Research Letters*, *16*(2), 024025.

Rehana, S. (2019). River water temperature modelling under Climate change using support vector regression. In S. K. Singh (Ed.), *Hydrology in a changing world*, *Springer Water Series* (pp. 171–183). Springer.

Richardson, J. (2019). Biological diversity in headwater streams. *Watermark*, *11*(2), 366.

RStudio Team. (2022). *RStudio: Integrated development environment for R*. RStudio PBC.

Rutherford, J. C., Marsh, N. A., Davies, P. M., & Bunn, S. E. (2004). Effects of patchy shade on stream water temperature: How quickly do small streams heat and cool? *Marine and Freshwater Research*, *55*(8), 737–748.

Simonyan, K., Vedaldi, A., & Zisserman, A. (2013). Deep inside convolutional networks: Visualising image classification models and saliency maps. arXiv preprint arXiv:1312.6034.

Smith, K. (1981). The prediction of river water temperatures / Prédiction des températures des eaux de rivière. *Hydrological Sciences Bulletin*, *26*(1), 19–32.

St-Hilaire, A., Morin, G., El-Jabi, N., & Caissie, D. (2000). Water temperature modelling in a small forested stream: Implication of forest canopy and soil temperature. *Canadian Journal of Civil Engineering*, *27*(6), 1095–1108.

Sundararajan, M., Taly, A., & Yan, Q. (2017). Axiomatic attribution for deep networks. In *International conference on machine learning* (pp. 3319–3328). PMLR.

Warnes, G. R., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M., & Venables, B. (2020). gplots: Various R Programming Tools for Plotting Data. R package version 3.1.1.

Webb, B., & Zhang, Y. (1997). Spatial and seasonal variability in the components of the river heat budget. *Hydrological Processes*, *11*(1), 79–101.

Webb, B., & Zhang, Y. (1999). Water temperatures and heat budgets in dorset chalk water courses. *Hydrological Processes*, *13*(3), 309–321.

Webb, B. W., Clack, P. D., & Walling, D. E. (2003). Water-air temperature relationships in a Devon river system and the role of flow. *Hydrological Processes*, *17*(15), 3069–3084.

Webb, B. W., Hannah, D. M., Moore, R. D., Brown, L. E., & Nobilis, F. (2008). Recent advances in stream and river temperature research. *Hydrological Processes: An International Journal*, *22*(7), 902–918.

Wild, R., Nagel, C., & Geist, J. (2023). *Climate change effects on hatching success and embryonic development of fish: Assessing multiple stressor responses in a large-scale mesocosm study* (164834). Science of The Total Environment.

Wondzell, S. M., Diabat, M., & Haggerty, R. (2019). What matters most: Are future stream temperatures more sensitive to changing air temperatures, discharge, or riparian vegetation? *JAWRA Journal of the American Water Resources Association*, *55*(1), 116–132.

Woodward, G., Perkins, D. M., & Brown, L. E. (2010). Climate change and freshwater ecosystems: Impacts across multiple levels of organization.

Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, *365*(1549), 2093–2106.

Zhu, S., Heddam, S., Nyarko, E. K., Hadzima-Nyarko, M., Piccolroaz, S., & Wu, S. (2019). Modeling daily water temperature for rivers: Comparison between adaptive neuro-fuzzy inference systems and artificial neural networks models. *Environmental Science and Pollution Research*, *26*(1), 402–420.

Zhu, S., Heddam, S., Wu, S., Dai, J., & Jia, B. (2019). Extreme learning machine-based prediction of daily water temperature for rivers. *Environmental Earth Sciences*, *78*(6), 1–17.

Zhu, S., Nyarko, E. K., Hadzima-Nyarko, M., Heddam, S., & Wu, S. (2019). Assessing the performance of a suite of machine learning models for daily river water temperature prediction. *PeerJ*, *7*, e7065.

Zurada, J. M., Malinowski, A., & Cloete, I. (1994). Sensitivity analysis for minimization of input data dimension for feedforward neural network. In *Proceedings of IEEE international symposium on circuits and systems, ISCAS…94* (Vol. 6, pp. 447–450). IEEE.

How to cite this article: Drainas, K., Kaule, L., Mohr, S., Uniyal, B., Wild, R., & Geist, J. (2023). Predicting stream water temperature with artificial neural networks based on open-access data. *Hydrological Processes*, *37*(10), e14991. https://doi.org/10.1002/hyp.14991

APPENDIX

A | Additional information on Materials and Methods

A.1 | Searching algorithms Scikit-learn

Our results were obtained using RandomizedSearch, which, given different options, optimized the architecture and hyperparameter combination of the ANNs for each individual input combination and waterbody.

RandomizedSearch certainly delivers a lower search quality than GridSearch, since it only determines local optima, unlike GridSearch, which delivers global optima. Still, as the ANN accuracy in our study demonstrated no weakness, in contrast to the results in the literature, we can support the use of RandomizedSearch, since it requires considerably less computing power and time. It is important to note, however, that even better results for the RMSE might be obtained with GridSearch and so it may be worth investing more time if fewer streams and less data needs to be processed or the time and capacity investment does not play a relevant role.

Difference between RandomizedSearch and GridSearch: While in RandomizedSearch, random sets of hyperparameters are used and tested, GridSearch tests all possible hyperparameter combinations systematically. The process can be accelerated by preselecting hyperparameters to reduce the total number of hyperparameter combinations. Of course, this again reduces the power of the search. In conclusion we suggest not using GridSearch if time and/or computing power are limited (see Figure 9).

A.2 | Results of RandomizedSearch

Using scikit-learn's RandomizedSearch, we determined an ANN with the hyperparameter combination leading to the lowest RMSE, for each waterbody and input combination. The RMSE values for all ANNs determined by RandomizedSearch are presented in Figure A1, according to waterbody. Figure A2 shows the same information but sorted by input combination. In Tables A3, A4, A5, and A6, these results are sorted by waterbody. The tables show which input combination for each stream reached what accuracy measures based on which hyperparameter combination. The abbreviations stand for the hyperparameters as indicated in the table below.

These combinations were attained by preselecting values for each hyperparameter based on prior experience. As stated above, preselection can reduce the power of the search, so we recommend including as many values as possible.

A.3 | Reliability of ANNs

Since common accuracy metrics only consider the differences between observed and predicted values, they are not suitable for assessing the reliability of the ANN, especially when it comes to changes in the database as expected for climate change scenarios. Hence, we also applied the reliability methodology as established in Mohr et al. (2021) on ANNs with the *allinputs* input combination, as determined by GridSearch.

A.3.1. | Perturbation analysis

Due to climate change, input variables will change in the future, e.g. air temperatures will rise. Since the training and testing datasets are retrieved from the past and cannot display future developments properly, a thorough analysis regarding changes in the input data is

Abbr.	Solver	Maximum iterations	Learning rate	Learning	Layers	Activation function
loc	lbfgs	100 000	0.0001	adaptive	5,20	logistic
rec	lbfgs	100 000	0.0001	constant	80,20,5	relu
tac	lbfgs	100 000	0.001	adaptive	80,10,5	tanh
werec	adam	100 000	0.01	constant	160,80,10	relu
wtac	adam	100 000	0.0001	adaptive	20 160,40	tanh
wtac2	lbfgs	100 000	0.0001	invscaling	80,10	tanh

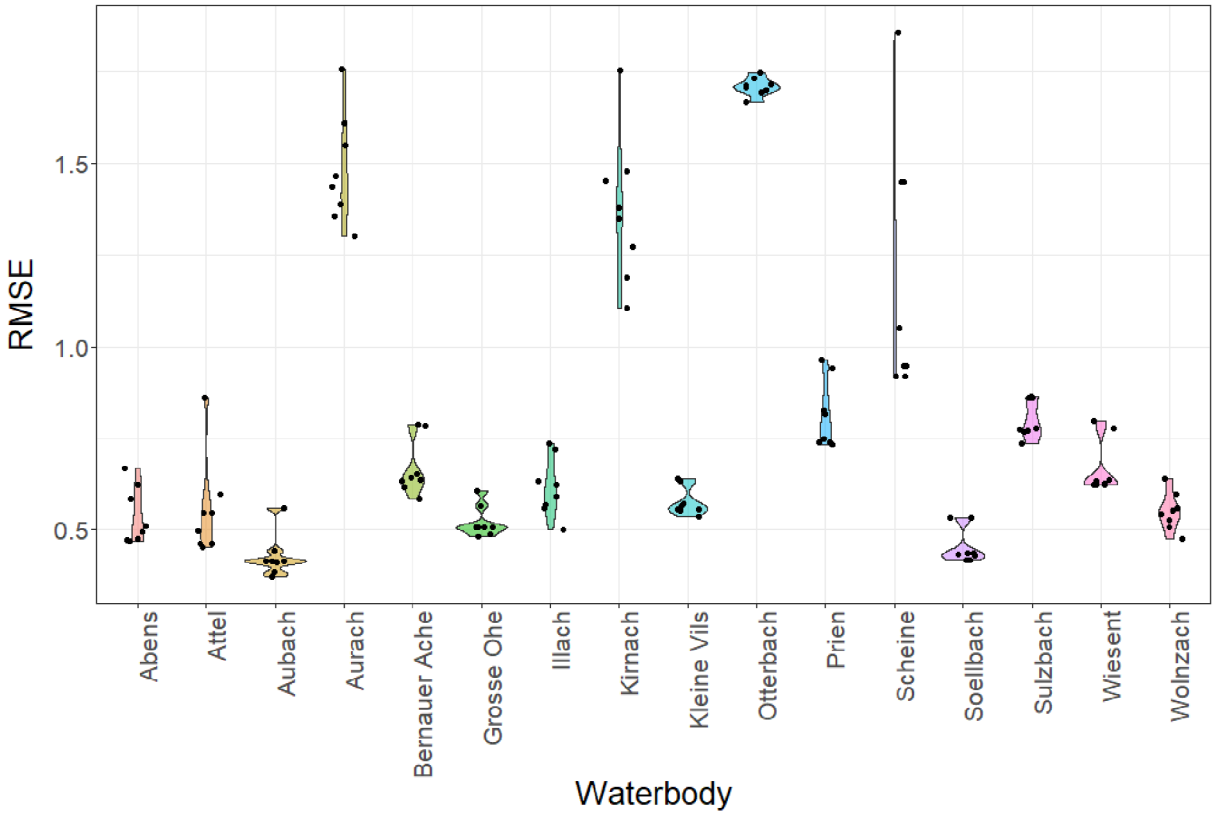

FIGURE A1 RMSE values for all input combinations per waterbody.

essential for the assessment of a model's reliability. In this study, we applied perturbation analysis to simulate changes in the input variables. For that, we perturbed every input value except the date by 0.01 (normalized) and evaluated the changes in the mean output. This reliability method is similar to accuracy metrics (comparison of observed vs. predicted data) but differs in that the input is changed.

A.3.2. | MinMax analysis

To consider how reasonable an ANN works regarding its predictions, it is useful to know the range of prediction values that the ANN can display. Therefore, we used MinMax analysis, where we chose random input values between 0 and 1 (normalized) to identify the operational range of each ANN. We optimized the initially chosen input and repeated the method 10 times for each ANN for the minimum and 10 times for the maximum value, respectively.

A.3.3. | Impact analysis

While the reaction of the ANN to perturbations and its operational range already give a good overview of its reliability, the so-called Impact Analysis, which is a method similar to *sensitivity analysis* (Zurada et al., 1994), can be used to determine which input the ANN is sensitive to, or, more specifically in our case, can be used to measure the importance of each input feature by determining its contribution to the water temperature calculation. With this information, it can be assessed whether single input parameters are weighted unreasonably high or low and hence predictions of future scenarios might not be reliable.

A.4 | Principal component analysis

To assess the environmental variables used to distinguish between the 16 assessed streams, we applied a principal component analysis (PCA) based on the normalized environmental variables that we compiled in the environmental dataset. The PCA and all subsequent multivariate analyses were calculated with the statistical software PRIMER v7 & PERMANOVA+ (Anderson et al., 2008).

A.5 | Correlation analysis

A.5.1. | Reliability metrics

As described for the accuracy metrics above, we also conducted a correlation analysis of the reliability metrics. To do this, we correlated all the environmental parameters of both resolutions (entire catchment

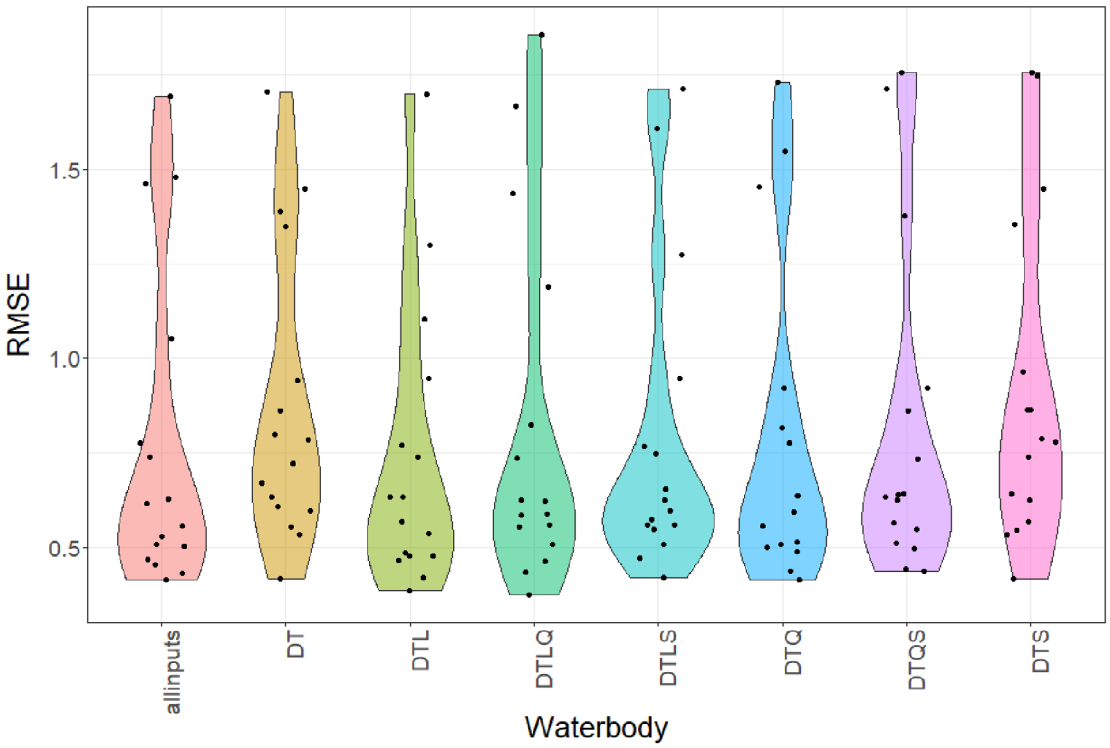

FIGURE A2 RMSE values for all streams per input combination. Input parameters: D, day of the year; DTQLS, allinputs; L, Water level, Q, discharge; S, sunshine duration; T, air temperature.

TABLE A1 Environmental parameter for distLM I.

Stream	Catchment	Longest river length	MW	HW	NW	MQ	HQ
Prien	52.82	15.15	0.24	220	3	1.70	128.00
Attel	66.14	11.01	0.28	188	10	1.00	24.70
Aubach	13.74	7.91	0.31	197	12	0.40	31.70
Söllbach	24.12	13.15	0.19	173	6	1.08	44.10
Bernauer Ache	36.53	8.80	0.48	215	29	0.80	37.10
Kleine Vils	43.25	9.55	0.51	229	25	0.97	55.90
Illach	25.09	19.99	0.50	225	34	0.79	24.90
Otterbach	91.74	22.79	0.89	235	70	0.83	32.10
Wiesent	135.38	14.46	1.34	195	110	1.05	6.88
Sulzbach	34.69	7.18	1.21	315	103	0.26	12.10
Abens	144.49	19.48	0.33	299	20	0.91	43.60
Aurach	123.59	28.20	1.37	278	110	0.66	20.20
Scheine	63.81	12.22	1.40	316	117	0.41	24.30
Große Ohe	18.70	5.15	0.44	165	19	0.60	23.80
Kirnach	25.31	21.15	0.35	223	10	0.77	49.50
Wolnzach	75.99	13.25	0.21	88	17	0.40	2.90

Note: Catchment, size of catchment (km^2); HQ, highest water discharge over entire period (m^3/s); HW, highest water level over entire period (cm); Longest river length, length of longest contributing river (km); MQ, mean water discharge over entire period (m^3/s); MW, mean water level over total period (m); NW, lowest water level over total period (cm); Stream, name of examined stream.

TABLE A2 Environmental parameter for distLM II.

Stream	Gauging station	DOD	IPO	Stations	Distances (km)
Abens	Mainburg	3782	32	02410 | **05404**	19.69 | 27.67
Attel	Assling	1821	40	01103 | **04261**	12.12 | 15.73
Aubach	Au	2593	40	**04261** | 03679	14.53 | 24.45
Aurach	Rothaurach	883	40	04280 | **03668**	4.56 | 29.13
BernauerAche	Bernau	4815	40	**00856** | 05941	13.90 | 17.13
GrosseOhe	Taferlruck	4169	40	**05800** | 01832	16.23 | 28.12
Illach	Engen	1904	32	00125 | **15 555**	12.17 | 28.56
Kirnach	Unterthingau	1012	40	**15 555** | 02559	12.73 | 14.15
KleineVils	Dietelskirchen	4355	40	13 710 | **03366**	10.76 | 26.78
Otterbach	Hammermuehle	8742	40	**04104** | 04559	10.99 | 30.10
Prien	Aschau	3505	40	05941 | **04261**	15.84 | 18.01
Scheine	Scheinfeld	996	36	01107 | 05149	21.55 | 21.82
Soellbach	BadWiessee	2162	36	02319 | 03679	20.06 | 33.67
Sulzbach	Koesfeld	4381	40	**00867** | 07428	3.25 | 15.31
Wiesent	Hollfeld	2379	40	**00320** | 00282	16.76 | 27.33
Wolnzach	Wolnzach	817	32	02410 | **05404**	13.56 | 22.86

Abbreviations: Distances (km), distances between GkD gauging station and DWD station; DOD, number indicating how many days served as data basis for training and testing; Gauging station, name of GkD gauging station from which water temperature, discharge and water level were obtained; IPO, maximum number of input values per output value; Stations, DWD stations from which air temperature data was used, bold indicates that sunshine duration was available (value from closer station preferred if possible); if no station is indicated in bold, no sunshine duration was available; Stream: name of stream investigated.

and 5 m riparian strip) with the mean perturbation values as well as with the results of the MinMax analysis that is, the minimum possible values and the maximum possible values for the ANN with the allinputs combination for each stream.

A.5.2. | Accuracy versus reliability metrics

Finally, we conducted a correlation analysis between the accuracy metrics RMSE, *R*, and PBIAS and the reliability metrics mean perturbation, minimum of MinMax analysis and maximum of MinMax analysis, to determine whether and to what extent the accuracy and reliability metrics agree with and/or complement each other.

B | Additional information on Results

B.1. | Assessment of ANNs

Impact analysis: The mean importance, as an indicator of the contribution of individual parameters for water temperature prediction, showed that the impact of individual input parameters strongly varied among streams (Figure B1). Calculating the mean over all streams, we observed that the water level of the current day (*L*) was the most important, with a value of 14%, followed by the mean air temperature of the closest DWD station of the current day (*mean_St*1), with a value of 11%. The greatest individual importance value of 35% was determined for the mean air temperature of the closest DWD station for the current day (*mean_St*1) at Grosse Ohe, for the water level of the current day (*L*) at Prien, and for the water level 3 days before the current day (*L*.3) at Sulzbach. On the other hand, we observed that for all streams, the sunshine duration (*S*) for the current and the previous days had no impact (0%), contradicting the findings of the accuracy metrics, in which some streams displayed the highest accuracy when *S* was included as input parameter.

MinMax analysis: MinMax analysis was applied to define the specific limits of water temperature prediction for each stream. For this analysis, values between 0 and 1 (normalized) were randomly recombined to identify the ANN's minimum and maximum water temperature predictions for each stream. The results of the MinMax analysis were in line with the above results, showing that the maximum and minimum range of the calculated values varied strongly depending on the stream's specifics.

Observed water temperature minima (0.34 ± 1.10°C) and maxima (19.55 ± 2.50°C) in the dataset differed from calculated minima (−11.96 ± 12.06°C) and maxima (74.05 ± 27.34°C). The mean delta (observed values minus calculated value) for the minimum values was 12.30 ± 11.88°C, with a maximum delta of 40.64°C (Sulzbach) and a minimum of 0.09°C (Kleine Vils). The mean delta for the maximum values was −54.50 ± 26.23°C with a maximum of −10.67°C (Wiesent) and a minimum of −106.89°C (Kirnach) (see Figure B2).

TABLE A3 Results of randomized search for Prien, Attel, Aubach and Soellbach, for explanations see Section A.2.

Waterbody	Inputs	Hyperparameter	RMSE	*R*	PBIAS	NSE
Prien	allinputs	rec	0.738	0.969	−0.111	0.938
Prien	DT	loc	0.940	0.949	−0.042	0.900
Prien	DTQ	loc	0.816	0.962	0.411	0.924
Prien	DTQS	rec	0.733	0.969	0.112	0.939
Prien	DTL	rec	0.737	0.969	0.237	0.938
Prien	DTLQ	wtac2	0.824	0.962	0.290	0.923
Prien	DTLS	rec	0.747	0.968	0.074	0.937
Prien	DTS	werec	0.962	0.946	−0.345	0.895
Mean			0.812	0.962	0.078	0.924
SD			0.087	0.009	0.227	0.017
Var			0.008	0.000	0.051	0.000
Attel	allinputs	rec	0.453	0.995	0.004	0.991
Attel	DT	loc	0.596	0.992	0.071	0.984
Attel	DTQ	tac	0.498	0.994	0.103	0.989
Attel	DTQS	rec	0.547	0.993	0.000	0.986
Attel	DTL	rec	0.464	0.995	0.108	0.990
Attel	DTLQ	tac	0.462	0.995	−0.127	0.990
Attel	DTLS	rec	0.546	0.993	0.158	0.986
Attel	DTS	tac	0.862	0.984	0.110	0.966
mean			0.554	0.993	0.053	0.985
SD			0.126	0.004	0.085	0.008
Var			0.016	0.000	0.007	0.000
Aubach	allinputs	rec	0.412	0.996	0.256	0.992
Aubach	DT	rec	0.415	0.996	0.314	0.992
Aubach	DTQ	tac	0.413	0.996	0.282	0.992
Aubach	DTQS	tac	0.443	0.995	0.272	0.991
Aubach	DTL	tac	0.385	0.997	0.163	0.993
Aubach	DTLQ	tac	0.373	0.997	0.033	0.994
Aubach	DTLS	loc	0.559	0.993	−0.066	0.986
Aubach	DTS	tac	0.416	0.996	0.424	0.992
mean			0.427	0.996	0.210	0.991
SD			0.054	0.001	0.149	0.002
Var			0.003	0.000	0.022	0.000
Soellbach	DT	rec	0.533	0.989	−0.183	0.977
Soellbach	DTQ	tac	0.437	0.992	−0.003	0.985
Soellbach	DTL	wtac2	0.419	0.993	−0.090	0.986
Soellbach	DTLQ	rec	0.433	0.992	−0.093	0.985
Mean			0.455	0.992	−0.079	0.983
SD			0.045	0.002	0.072	0.003
Var			0.002	0.000	0.005	0.000

TABLE A4 Results of randomized search for Bernauer Ache, Kleine Vils, Illach and Otterbach, for explanations see Section A.2.

Waterbody	Inputs	Hyperparameter	RMSE	*R*	PBIAS	NSE
Bernauer Ache	allinputs	loc	0.616	0.986	−0.350	0.973
Bernauer Ache	DT	tac	0.784	0.978	−0.731	0.956
Bernauer Ache	DTQ	tac	0.637	0.986	−0.378	0.971
Bernauer Ache	DTQS	rec	0.641	0.985	−0.310	0.971
Bernauer Ache	DTL	loc	0.634	0.986	−0.458	0.971
Bernauer Ache	DTLQ	tac	0.586	0.988	−0.315	0.976
Bernauer Ache	DTLS	rec	0.653	0.985	−0.562	0.970
Bernauer Ache	DTS	loc	0.787	0.978	−0.570	0.956
mean			0.667	0.984	−0.459	0.968
SD			0.071	0.004	0.140	0.007
Var			0.005	0.000	0.020	0.000
Kleine Vils	allinputs	loc	0.555	0.997	0.152	0.993
Kleine Vils	DT	tac	0.632	0.996	−0.077	0.991
Kleine Vils	DTQ	rec	0.555	0.997	−0.025	0.993
Kleine Vils	DTQS	rec	0.564	0.997	−0.081	0.993
Kleine Vils	DTL	tac	0.535	0.997	−0.105	0.994
Kleine Vils	DTLQ	tac	0.552	0.997	−0.126	0.993
Kleine Vils	DTLS	tac	0.572	0.997	−0.131	0.993
Kleine Vils	DTS	tac	0.640	0.996	−0.178	0.991
mean			0.576	0.996	−0.071	0.993
SD			0.036	0.000	0.094	0.001
Var			0.001	0.000	0.009	0.000
Illach	allinputs	loc	0.503	0.994	−0.136	0.989
Illach	DT	loc	0.721	0.988	−0.559	0.977
Illach	DTQ	rec	0.592	0.992	−0.659	0.984
Illach	DTQS	rec	0.633	0.991	−0.470	0.982
Illach	DTL	tac	0.567	0.993	−0.026	0.986
Illach	DTLQ	rec	0.622	0.991	−0.091	0.983
Illach	DTLS	rec	0.560	0.993	−0.310	0.986
Illach	DTS	rec	0.737	0.988	−0.967	0.976
mean			0.617	0.991	−0.402	0.983
SD			0.075	0.002	0.302	0.004
Var			0.006	0.000	0.091	0.000
Otterbach	allinputs	rec	1.693	0.957	−0.254	0.915
Otterbach	DT	rec	1.704	0.956	−0.295	0.914
Otterbach	DTQ	werec	1.730	0.955	−1.134	0.912
Otterbach	DTQS	werec	1.713	0.956	−0.889	0.913
Otterbach	DTL	rec	1.700	0.956	−0.160	0.915
Otterbach	DTLQ	rec	1.667	0.958	−0.248	0.918
Otterbach	DTLS	werec	1.714	0.956	0.069	0.913
Otterbach	DTS	werec	1.747	0.954	0.369	0.910
mean			1.708	0.956	−0.318	0.914
SD			0.023	0.001	0.454	0.002
Var			0.001	0.000	0.206	0.000

TABLE A5 Results of randomized search for Wiesent, Sulzbach, Abens and Aurach, for explanations see Section A.2.

Waterbody	Inputs	Hyperparameter	RMSE	*R*	PBIAS	NSE
Wiesent	allinputs	rec	0.626	0.985	−0.060	0.970
Wiesent	DT	rec	0.798	0.976	0.076	0.951
Wiesent	DTQ	tac	0.637	0.985	0.077	0.969
Wiesent	DTQS	rec	0.623	0.985	0.158	0.970
Wiesent	DTL	tac	0.633	0.985	0.297	0.969
Wiesent	DTLQ	tac	0.625	0.985	0.153	0.970
Wiesent	DTLS	rec	0.624	0.985	−0.003	0.970
Wiesent	DTS	rec	0.778	0.977	−0.249	0.954
mean			0.668	0.983	0.056	0.965
SD			0.070	0.004	0.154	0.008
Var			0.005	0.000	0.024	0.000
Sulzbach	allinputs	rec	0.774	0.989	−0.176	0.978
Sulzbach	DT	rec	0.861	0.986	−0.380	0.972
Sulzbach	DTQ	rec	0.776	0.989	−0.309	0.977
Sulzbach	DTQS	tac	0.861	0.986	−0.424	0.972
Sulzbach	DTL	tac	0.770	0.989	−0.320	0.978
Sulzbach	DTLQ	tac	0.736	0.990	−0.286	0.980
Sulzbach	DTLS	rec	0.766	0.989	−0.310	0.978
Sulzbach	DTS	loc	0.863	0.986	−0.380	0.972
mean			0.801	0.988	−0.323	0.976
SD			0.049	0.001	0.071	0.003
Var			0.002	0.000	0.005	0.000
Abens	allinputs	tac	0.468	0.995	−0.084	0.990
Abens	DT	loc	0.670	0.990	0.081	0.980
Abens	DTQ	wtac2	0.512	0.994	0.094	0.988
Abens	DTQS	rec	0.496	0.994	−0.040	0.989
Abens	DTL	wtac2	0.474	0.995	0.135	0.990
Abens	DTLQ	rec	0.585	0.992	−0.045	0.984
Abens	DTLS	tac	0.471	0.995	−0.046	0.990
Abens	DTS	rec	0.623	0.991	0.045	0.982
mean			0.537	0.993	0.017	0.987
SD			0.073	0.002	0.076	0.004
Var			0.005	0.000	0.006	0.000
Aurach	allinputs	rec	1.464	0.961	0.507	0.924
Aurach	DT	rec	1.388	0.965	−0.082	0.931
Aurach	DTQ	rec	1.549	0.957	1.075	0.914
Aurach	DTQS	rec	1.756	0.944	0.877	0.890
Aurach	DTL	rec	1.301	0.969	−0.113	0.940
Aurach	DTLQ	rec	1.436	0.963	0.498	0.926
Aurach	DTLS	rec	1.609	0.954	0.912	0.908
Aurach	DTS	rec	1.355	0.967	0.632	0.934
mean			1.482	0.960	0.538	0.921
SD			0.140	0.008	0.413	0.015
Var			0.019	0.000	0.170	0.000

TABLE A6 Results of randomized search for Scheine, Grosse Ohe, Kirnach and Wolnzach, for explanations see Section A.2.

Waterbody	Inputs	Hyperparameter	RMSE	*R*	PBIAS	NSE
Scheine	DT	rec	1.449	0.949	0.948	0.900
Scheine	DTQ	rec	0.920	0.980	−0.676	0.960
Scheine	DTL	rec	0.947	0.978	−0.078	0.957
Scheine	DTLQ	rec	1.857	0.921	−1.687	0.836
mean			1.193	0.964	−0.191	0.927
SD			0.328	0.020	0.816	0.042
Var			0.108	0.000	0.666	0.002
Grosse Ohe	allinputs	rec	0.506	0.993	−0.440	0.987
Grosse Ohe	DT	tac	0.607	0.990	−0.603	0.981
Grosse Ohe	DTQ	wtac2	0.488	0.994	−0.517	0.988
Grosse Ohe	DTQS	loc	0.509	0.993	−0.561	0.986
Grosse Ohe	DTL	tac	0.483	0.994	−0.344	0.988
Grosse Ohe	DTLQ	tac	0.508	0.993	−0.297	0.987
Grosse Ohe	DTLS	rec	0.508	0.993	−0.309	0.986
Grosse Ohe	DTS	loc	0.566	0.992	−0.651	0.983
mean			0.522	0.993	−0.465	0.986
SD			0.040	0.001	0.129	0.002
Var			0.002	0.000	0.017	0.000
Kirnach	allinputs	rec	1.479	0.964	−1.794	0.926
Kirnach	DT	rec	1.348	0.969	0.074	0.939
Kirnach	DTQ	rec	1.453	0.964	−0.792	0.929
Kirnach	DTQS	rec	1.378	0.968	−1.317	0.936
Kirnach	DTL	rec	1.104	0.979	−0.767	0.959
Kirnach	DTLQ	loc	1.188	0.976	−1.093	0.952
Kirnach	DTLS	rec	1.273	0.972	−0.184	0.945
Kirnach	DTS	werec	1.755	0.947	−0.482	0.896
mean			1.372	0.967	−0.794	0.935
SD			0.187	0.009	0.569	0.018
Var			0.035	0.000	0.323	0.000
Wolnzach	allinputs	rec	0.528	0.985	0.426	0.970
Wolnzach	DT	rec	0.554	0.983	0.182	0.967
Wolnzach	DTQ	loc	0.508	0.987	0.659	0.972
Wolnzach	DTQS	rec	0.638	0.979	0.066	0.956
Wolnzach	DTL	rec	0.476	0.988	−0.048	0.976
Wolnzach	DTLQ	rec	0.558	0.983	0.005	0.966
Wolnzach	DTLS	rec	0.596	0.981	0.778	0.962
Wolnzach	DTS	rec	0.543	0.984	0.305	0.968
mean			0.550	0.984	0.297	0.967
SD			0.047	0.003	0.285	0.006
Var			0.002	0.000	0.081	0.000

Perturbation analysis: We applied perturbation analysis to test the degree to which the ANNs' predictions changed when historical input data varied by 0.01 (normalized). The mean perturbation value over all streams was 2.620 ± 2.109°C, with the highest mean perturbation observed in Otterbach (9.981°C) and the lowest in Wolnzach (0.985°C).

B.2. | Relationship between accuracy and reliability metrics

Correlation analysis (see Table B1) resulted in one highly significant correlation between RMSE and mean perturbation ($\rho = 0.853$, $p < 0.001$) and three significant correlations: between *R* and mean perturbation ($\rho = -0.599$, $p < 0.05$), between RMSE and MinMax-max ($\rho = 0.538$, $p < 0.05$) and between PBIAS and MinMax-max ($\rho = -0.582$, $p < 0.05$). There were no significant correlations between any of the accuracy metrics and MinMax-Min.

B.3. | Environmental characteristics of sites

The PCA of environmental conditions across the streams (Figure B3) showed that the 16 sites were broadly distributed along multiple environmental gradients. The first PC axis, covering 26.8% of the observed variation (Eigenvalue = 6.44), structured streams primarily according to the proportion of natural and forested vegetation and water bodies in their surroundings. It exemplifies that the streams Grosse Ohe, Soellbach and Bernauer Ache feature a higher share of natural vegetation than such streams as the Scheine, Kleine Vils or Sulzbach. Also hydrological features of the streams investigated, such as NW and HW, were reflected by PC1, with streams in the negative space of PC1 tending to have higher mean and high water levels than those in the positive space. The second PC axis, making up 16.6% of the observed variation in the data set (Eigenvalue = 3.98), grouped streams largely according to the proportion of agricultural and urban land use (with a high share, for example, along Sulzbach and Wolnzach and a low share in Kirnach, Prien and Illach), while the proportion of grassland in the surroundings and the total length of the river upstream from the sampling site grouped streams in the opposite direction. Detailed proportions of land use are depicted in Table C1. Further information on environmental parameters is depicted in Tables A1 and A2.

B.4. | Environmental predictors of ANN reliability metrics

Regarding the overall catchment resolution (Table B2 top), we determined a significantly positive correlation between mean perturbation

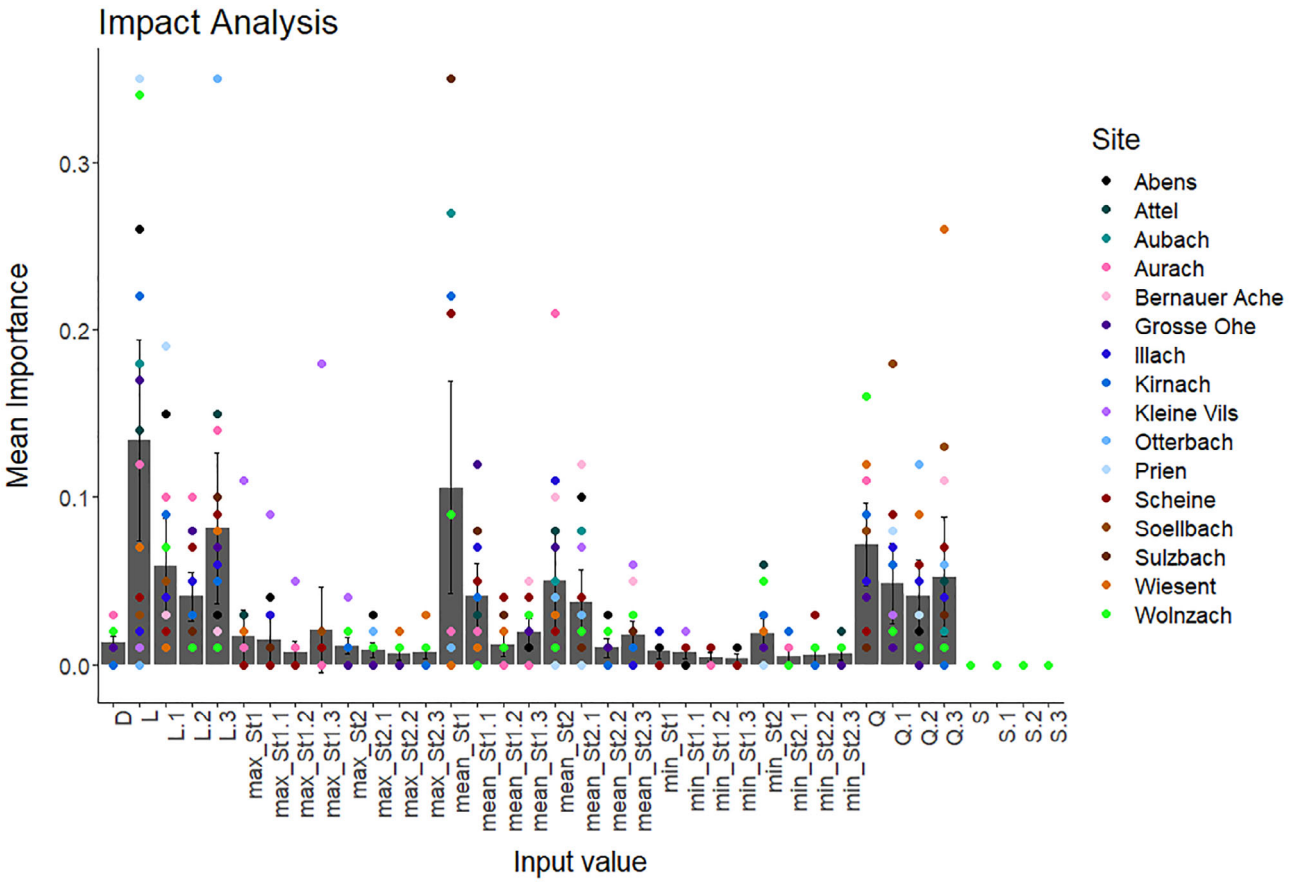

FIGURE B1 Impact analysis for all input parameters in each stream. Inputs on *x*-axis indexed as below. Whiskers mark 95% confidence intervals and bars mark mean importance for each input. St indicates the station from which air temperature was received (St1 = DWD station closest to GkD gauging station, St2 = DWD station second-closest to GkD gauging station). The "addendum.No" indicates how many days prior to *D* the data is from (0.1 = the day before D, 0.2 = 2 days before D, 0.3 = 3 days before D). Input values: *D*, day of the year; *L*, water level; max, maximum air temperature; mean, mean air temperature; min, minimum air temperature; *Q*, discharge; *S*, sunshine duration.

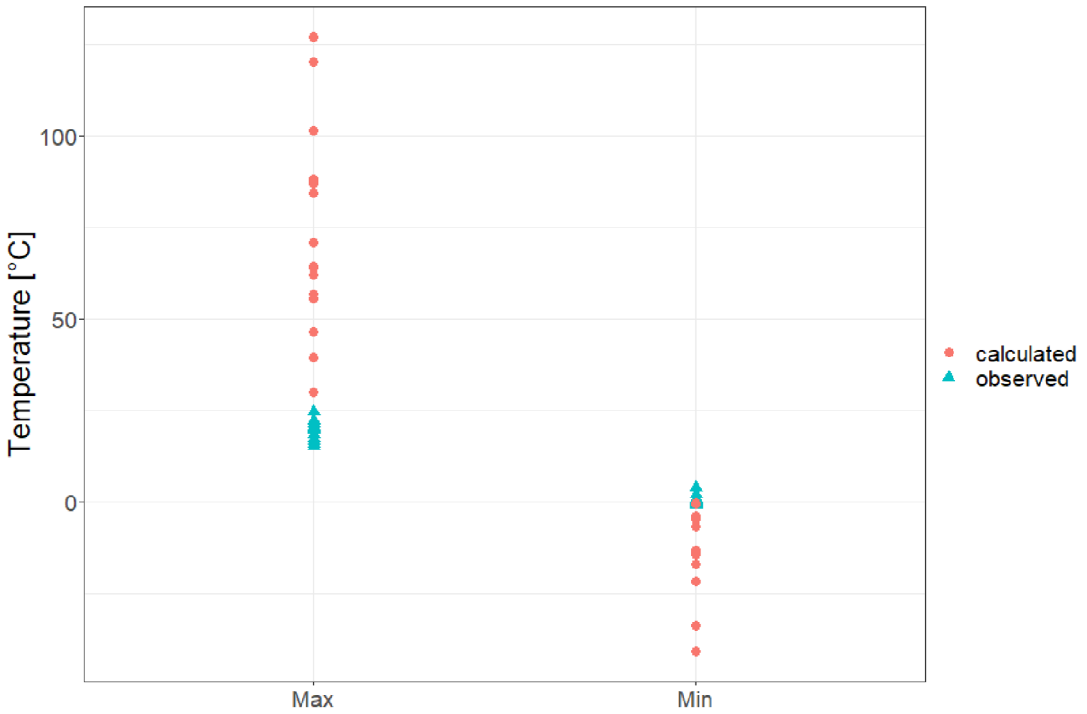

FIGURE B2 Comparison of calculated and observed minimum and maximum values for all waterbodies. Calculated values were determined by MinMax analysis, observed values were retrieved from the datasets.

TABLE B1 Correlations between evaluation and assessment metrics.

	RMSE	*R*	PBIAS
Mean perturbation	0.853***	−0.600*	−0.185
MinMax_min	−0.053	0.157	−0.091
MinMax_max	0.538*	−0.213	−0.582*

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

and HW ($r = 0.67$, $p < 0.01$) as well as a significantly negative relationship between the minimum values of the MinMax analysis and DOD ($r = -0.60$, $p < 0.05$) and between the maximum values of the MinMax analysis and semi-natural ($r = -0.55$, $p < 0.05$). For the 5 m riparian strip resolution (see Table B2 bottom), there was a significantly positive correlation between mean perturbation and grassland ($r = 0.52$, $p < 0.05$) as well as between the minimum values of the MinMax analysis and grassland ($r = 0.50$, $p < 0.05$).

C. | Additional information on the Discussion

C.1. | Accuracy and reliability

Not all accuracy vs. reliability metrics correlated significantly. This confirmed the finding of Mohr et al. (2021), that the use of accuracy metrics alone is not sufficient and must be supplemented with reliability metrics. Still, we can conclude that as accuracy metric, the RMSE is the most suitable one of those we used to reflect the reliability of an ANN. We are able to conclude this thanks to the significant correlations both to mean perturbation and to the maximum values obtained in the MinMax analysis, while only one significant correlation was demonstrated for R and PBIAS, respectively. We also observed that the RMSE had a greater resolution and hence contributed more significant relationships with environmental parameters than R, probably because it had a higher potential to reflect the high-resolution dynamics of hydrologic parameters. This further increased the benefit of the RMSE and confirms the plausibility of its frequent use for measuring the accuracy of water temperature prediction with ANNs (Ahmadi-Nedushan et al., 2007; Caissie et al., 1998; Chenard & Caissie, 2008; Cho & Lee, 2012; Feigl et al., 2021; Graf et al., 2019; Hadzima-Nyarko et al., 2014; Qiu et al., 2020; Quan et al., 2020; Rabi et al., 2015; Rahmani et al., 2020; Rehana, 2019; St-Hilaire et al., 2000; Zhu, Nyarko, Hadzima-Nyarko, et al., 2019).

In this study, we additionally employed PBIAS as an accuracy metric, which is unusual for water temperature prediction with ANNs. Although we see advantages in combining different accuracy metrics and including the PBIAS due to the different aspects of model performance it highlights, in this study we were not able to find any general correlations between the environmental parameters

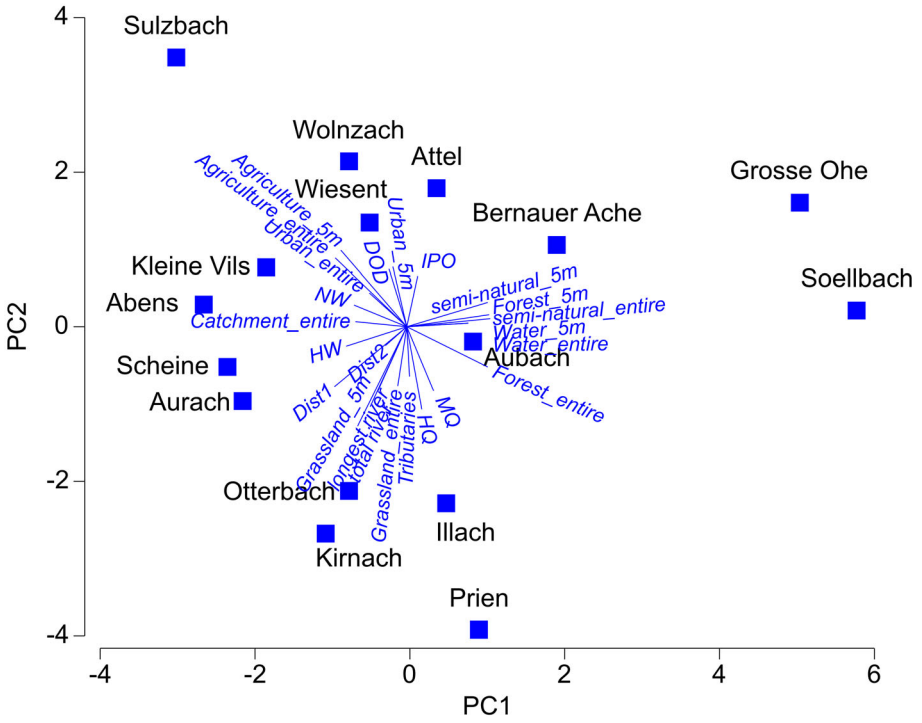

FIGURE B3 Principal component analysis plot, all stream sites broadly distributed along multiple environmental gradients. Streams structured by PC axis 1 mainly according to the proportion of natural and forested vegetation and waterbodies in their surroundings, as well as by NW and HW. Structure by PC axis 2 mainly according to proportion of agricultural and urban land use.

TABLE B2 Correlation analysis of environmental parameters versus robustness measures, entire catchment and 5 m riparian strip.

	Perturbation	Min	Max
Entire catchment			
Total river length	0.406	−0.141	0.191
Longest river length	0.344	0.238	0.413
Agriculture	0.020	0.065	−0.128
Forest	0.003	0.000	−0.178
Grassland	0.456	0.026	0.426
Semi-natural	−0.328	−0.081	−0.546*
Urban	0.068	0.135	−0.074
Water	−0.140	0.084	−0.028
Catchment	0.044	0.221	−0.326
MW	0.421	−0.259	0.459
HW	0.671**	−0.191	0.495
NW	0.250	−0.225	0.268
MQ	−0.041	0.321	−0.289
HQ	0.279	0.115	0.147
DOD	0.038	−0.588*	−0.171
IPO	−0.050	−0.442	−0.070
Dist1	0.397	0.265	0.408
Dist2	0.121	0.024	0.283
Tributaries	0.324	−0.112	0.094
5 m riparian strip			
Total river length	0.406	−0.141	0.191
Longest river length	0.344	0.238	0.203
Agriculture	0.037	0.197	−0.009

TABLE B2 (Continued)

	Perturbation	Min	Max
Forest	−0.421	−0.026	−0.309
Grassland	0.521*	0.001	0.498*
Semi-natural	−0.365	−0.091	−0.220
Urban	−0.162	0.110	−0.415
Water	−0.140	0.084	−0.028
Area	0.341	−0.074	0.097

Abbreviations: Agriculture, forest, grassland, semi-natural, urban, water: land use; Area, total area of riparian strip; Area, total buffer area; Catchment, Total size of all contributing catchments; D, Day of the year; Dist1, distance between GkD station and DWD station 1; Dist2, distance between GkD station and DWD station 2; DOD, number of days for which data was used; DTQLS, allinputs; HQ, highest measured discharge; HW, highest measured water level; IPO, number of input data points per output data point; *L*, water level; Longest river length, the length of the longest contributing river; Max, maximum determined by MinMax-analysis; Min, minimum determined by MinMax-analysis; MQ, mean discharge; MW, mean water level; NW, lowest measured water level; Perturbation, mean perturbation determined by perturbation analysis; *Q*, discharge; *S*, sunshine duration; T, air temperature; Total river length, sum of lengths of all contributing rivers; Tributaries, number of tributaries.
*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

and the PBIAS. This might be because the PBIAS reflects variation in two directions, but the direction of estimation (over- or underestimation) does not necessarily correlate with an environmental parameter in only one direction. Even though the PBIAS showed no consistent trends in the correlation analysis with either environmental parameters or reliability metrics, the significant correlation between the PBIAS and the maximum values of the MinMax analysis showed that with increasing MinMax-max values, the ANNs tended to overestimate water temperature. This overestimation was pronounced for Kirnach, a stream with a very high proportion of grassland (72.58%) and a very low proportion of semi-natural land cover (0.01%). In contrast, underestimation of water temperature was pronounced for Aurach, a long stream with a large catchment. These findings were also confirmed by the DistLM analysis of environmental predictors of evaluation metrics, in which PBIAS/overestimation was associated with high proportions of grassland, particularly in the 5 m riparian strip. Consequently, it would be advisable to carefully check for both over- and underestimation of the water temperature prediction, particularly in catchments with high proportions of open-canopy landscape (Figures C1–C6).

TABLE C1 Land use in entire catchment and in 5 m riparian strip.

Stream	Total river length	Agriculture	Forest	Grassland	Semi-natural	Urban	Water
Prien	15.15	x	70.91	12.03	14.98	2.09	x
Attel	11.01	21.03	30.81	37.74	1.04	9.38	x
Aubach	7.91	x	34.30	63.92	x	1.79	x
Soellbach	13.15	x	77.92	1.62	18.58	1.83	0.05
Bernauer Ache	8.80	0.08	46.55	23.23	25.37	4.77	x
Kleine Vils	9.55	72.68	18.43	6.01	x	2.88	x
Illach	19.99	x	33.04	61.96	3.32	1.68	x
Otterbach	41.43	23.89	41.12	33.17	0.31	1.52	x
Wiesent	26.13	56.25	37.62	3.27	0.71	2.15	x
Sulzbach	13.89	75.33	11.55	10.12	x	2.99	x
Abens	19.48	62.78	21.59	8.45	0.81	6.38	x
Aurach	28.20	40.06	44.56	10.26	x	5.12	x
Scheine	12.22	38.62	41.44	17.04	x	2.89	x
Große Ohe	11.95	x	69.22	x	30.78	x	x
Kirnach	39.51	x	23.91	72.58	0.01	3.50	x
Wolnzach	13.25	62.06	24.52	7.47	1.43	4.51	x
Stream	Mean river length	Agriculture	Forest	Grassland	Semi-natural	Urban	Water
Prien	15.15	x	25.06	66.77	x	8.18	x
Attel	11.01	24.65	45.09	x	8.75	21.51	x
Aubach	7.91	x	58.12	34.00	x	7.88	x
Soellbach	13.15	x	73.06	1.81	12.62	11.33	1.17
Bernauer Ache	8.80	x	57.02	19.18	6.68	17.12	x
Kleine Vils	9.55	41.76	2.14	52.30	x	3.79	x
Illach	19.99	x	33.26	54.14	12.60	x	
Otterbach	13.81	14.44	44.64	44.72	x	3.05	x
Wiesent	13.07	8.07	62.19	17.77	6.49	8.74	x
Sulzbach	6.95	67.38	x	21.31	x	11.31	x
Abens	19.48	32.37	6.86	44.65	x	16.12	x
Aurach	28.20	16.67	45.02	29.06	x	9.25	x
Scheine	12.22	24.94	11.21	55.46	x	8.39	x
Große Ohe	3.98	x	73.46	x	26.54	x	x
Kirnach	19.75	x	21.32	70.58	x	8.1	x
Wolnzach	13.25	39.14	32.86	8.16	x	19.84	x

Note: Top: land use in entire catchment. Bottom: land use in 5 m riparian strip for whole river.
Abbreviations: Agriculture, forest, grassland, semi-natural, urban, water: Proportion of land use in percent, for 5 m riparian strip as mean over all arms; Mean river length: (for 5 m riparian strips) If stream contained more than one arm, this is the mean of the lengths of the arms in km; Stream, name of stream investigated; Total river length, sum of lengths of all contributing rivers in km.

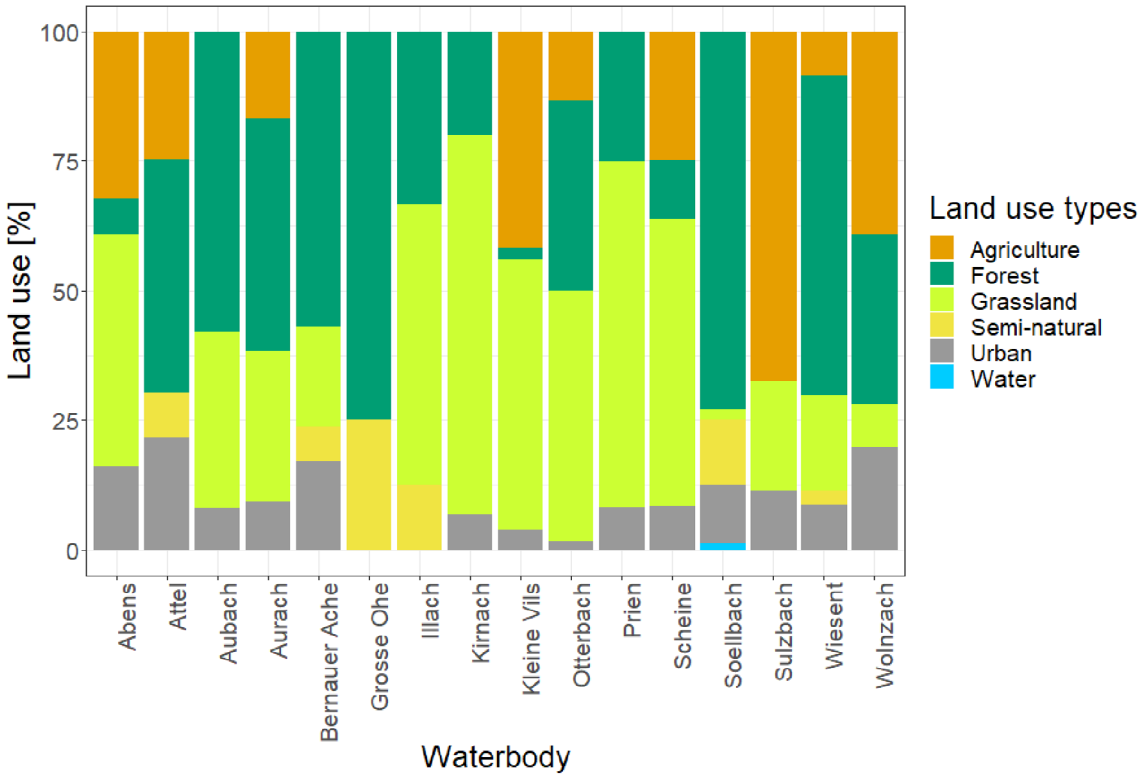

FIGURE C1 Cumulative barplot illustrating the shares of land use in the 5 m riparian strip.

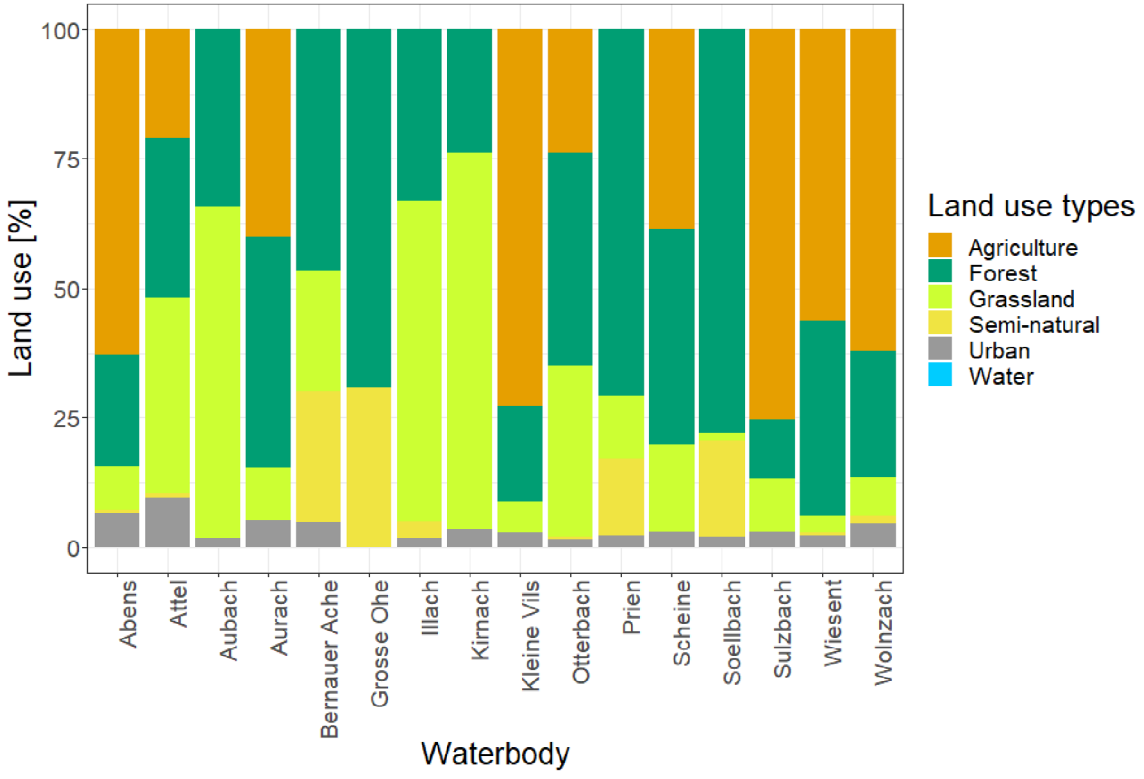

FIGURE C2 Cumulative barplot illustrating the shares of land use in the entire catchment.

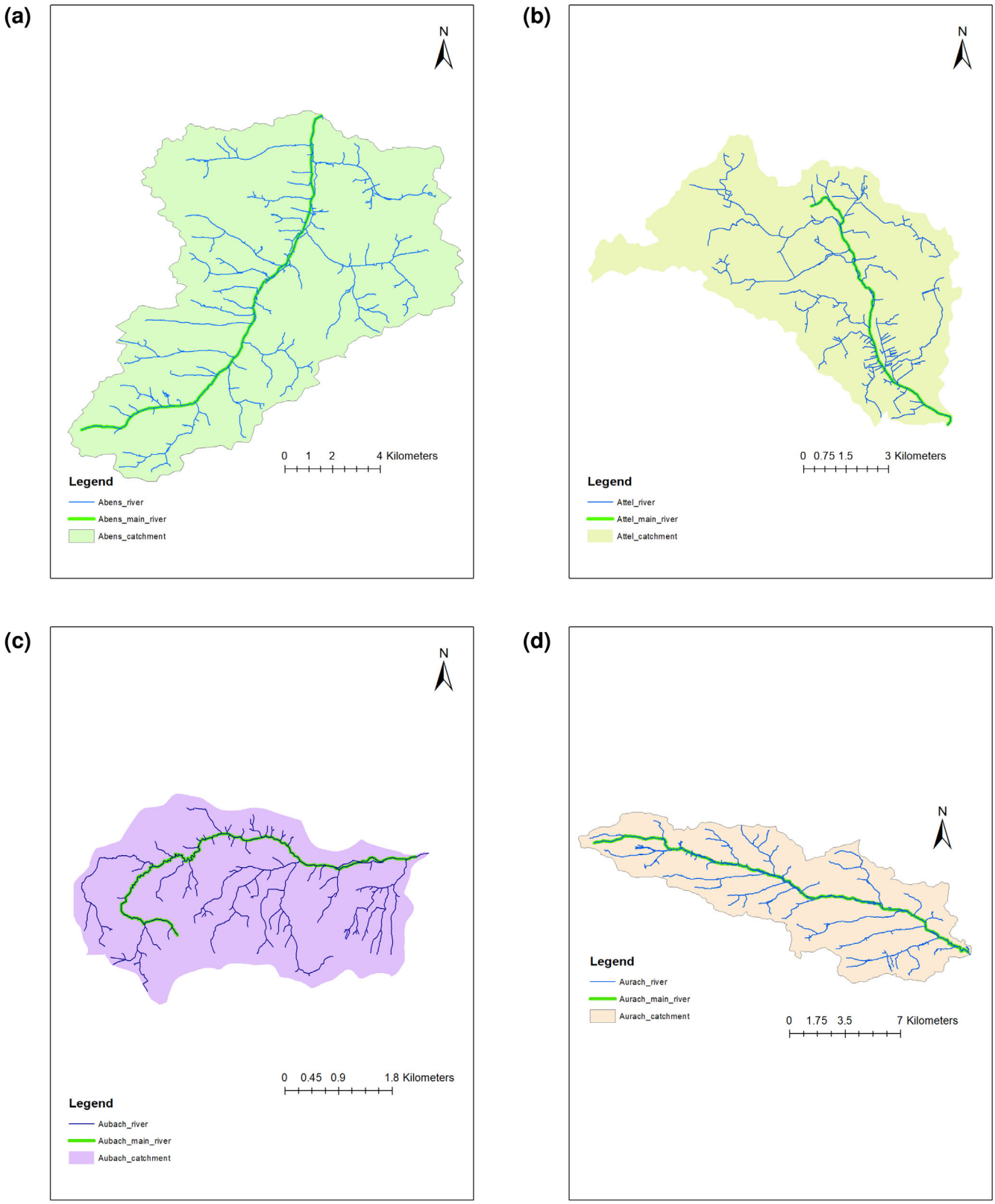

FIGURE C3 Catchments (a) Abens, (b) Attel, (c) Aubach, and (d) Aurach.

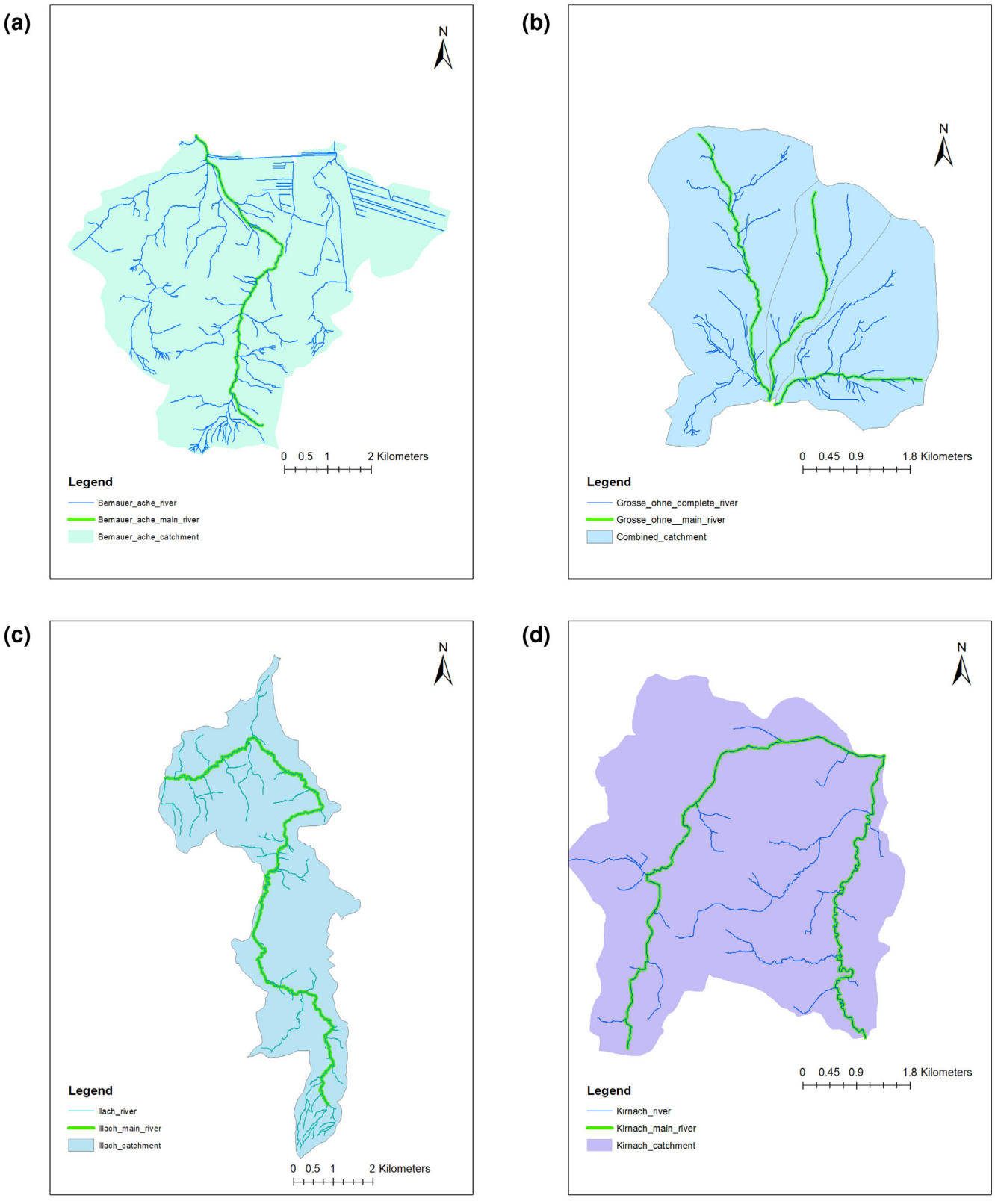

FIGURE C4 Catchments (a) Bernauer Ache, (b) Grosse Ohe, (c) Illach, and (d) Kirnach.

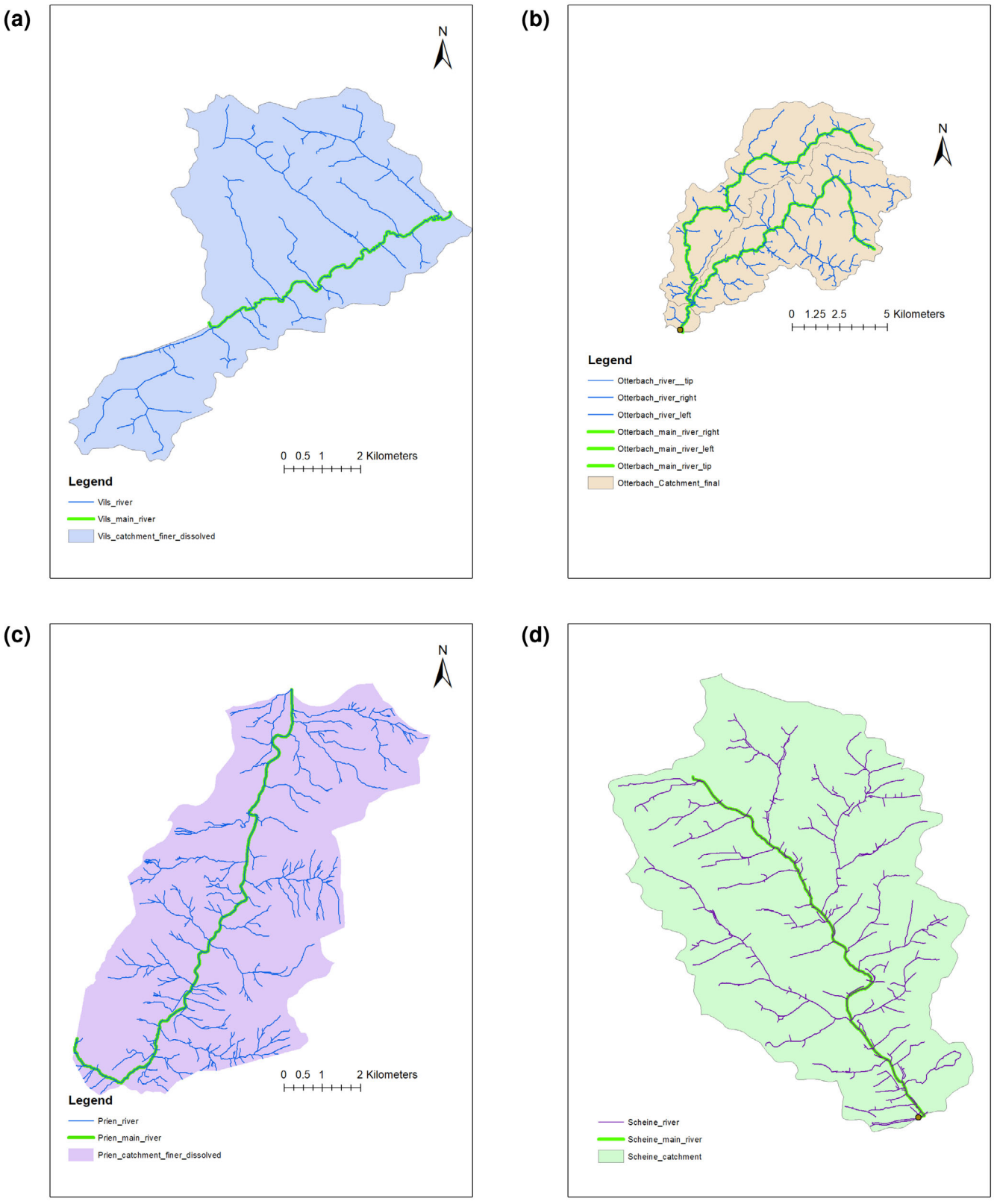

FIGURE C5 Catchments (a) Kleine Vils, (b) Otterbach, (c) Prien, and (d) Scheine.

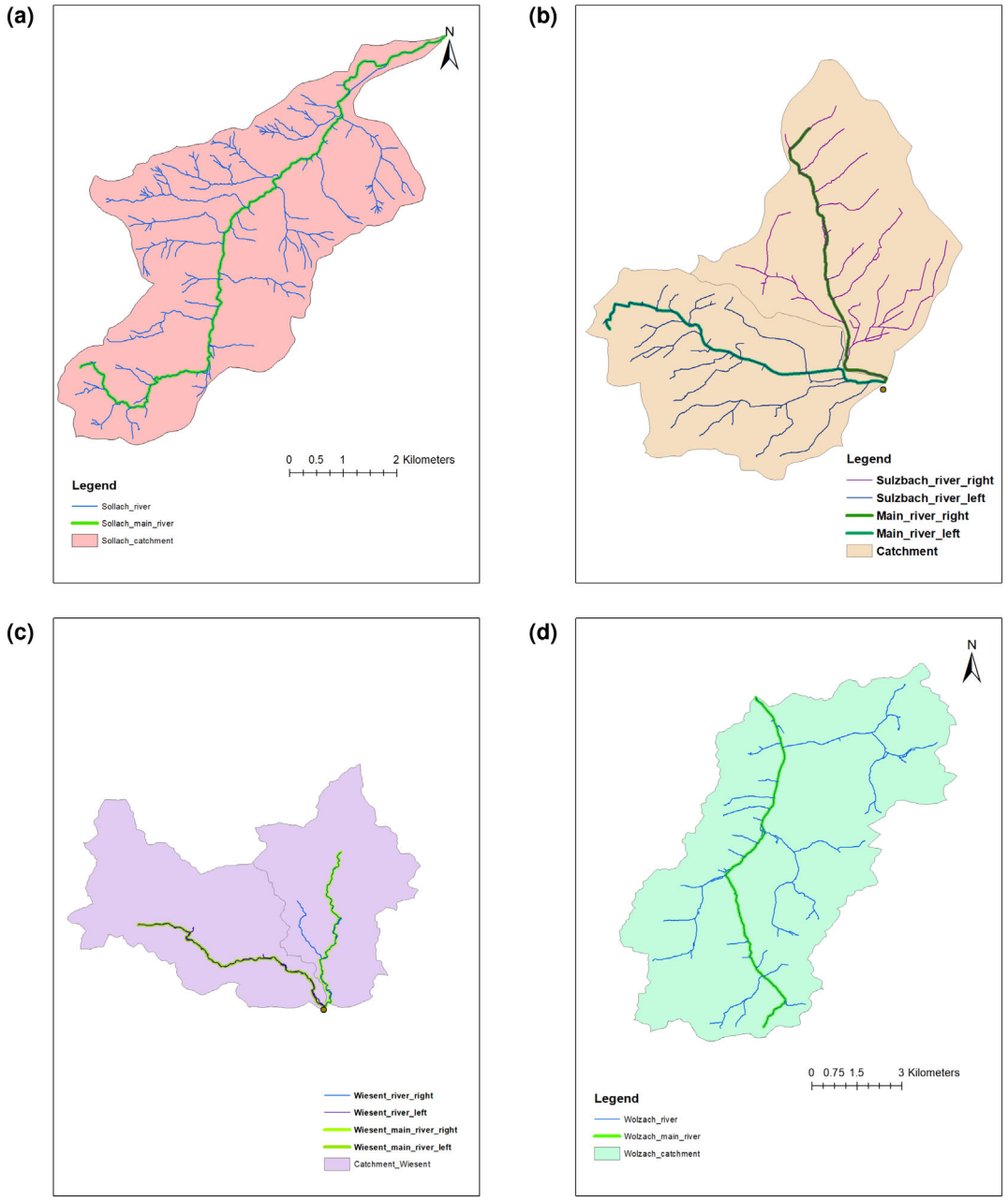

FIGURE C6 Catchments (a) Soellbach, (b) Sulzbach, (c) Wiesent, and (d) Wolnzach.