PHILOSOPHICAL TRANSACTIONS B

rstb.royalsocietypublishing.org

Research

Cite this article: Le TDH, Kattwinkel M, Schützenmeister K, Olson JR, Hawkins CP, Schäfer RB. 2019 Predicting current and future background ion concentrations in German surface water under climate change. *Phil. Trans. R. Soc. B* **374**: 20180004.
http://dx.doi.org/10.1098/rstb.2018.0004

Accepted: 20 September 2018

One contribution of 23 to a theme issue 'Salt in freshwaters: causes, ecological consequences and future prospects'.

Subject Areas:
environmental science

Keywords:
salinization, streams chemistry, modelling, forecast

Author for correspondence:
Trong Dieu Hien Le
e-mail: dieuhien@uni-landau.de

Electronic supplementary material is available online at https://dx.doi.org/10.6084/m9.figshare.c.4269398.

THE ROYAL SOCIETY PUBLISHING

Predicting current and future background ion concentrations in German surface water under climate change

Trong Dieu Hien Le[1,2], Mira Kattwinkel[1], Klaus Schützenmeister[1], John R. Olson[3], Charles P. Hawkins[4] and Ralf B. Schäfer[1]

[1]Institute for Environmental Sciences, University Koblenz-Landau, Fortstraße 7, 76829 Landau, Germany
[2]Faculty of Resources and Environment, University of Thu Dau Mot, 06 Tran Van On street, Thu Dau Mot City, Binh Duong 820000, Vietnam
[3]School of Natural Sciences, California State University Monterey Bay, Seaside, CA 93955, USA
[4]Watershed Science, Utah State University, Logan, UT 84322, USA

LTDH, 0000-0002-9749-3416; JRO, 0000-0003-1456-6681

Salinization of surface waters is a global environmental issue that can pose a regional risk to freshwater organisms, potentially leading to high environmental and economic costs. Global environmental change including climate and land use change can increase the transport of ions into surface waters. We fit both multiple linear regression (LR) and random forest (RF) models on a large spatial dataset to predict Ca^{2+} (266 sites), Mg^{2+} (266 sites), and SO_4^{2-} (357 sites) ion concentrations as well as electrical conductivity (EC—a proxy for total dissolved solids with 410 sites) in German running water bodies. Predictions in both types of models were driven by the major factors controlling salinity including geologic and soil properties, climate, vegetation and topography. The predictive power of the two types of models was very similar, with RF explaining 71–76% of the spatial variation in ion concentrations and LR explaining 70–75% of the variance. Mean squared errors for predictions were all smaller than 0.06. The factors most strongly associated with stream ion concentrations varied among models but rock chemistry and climate were the most dominant. The RF model was subsequently used to forecast the changes in EC that were likely to occur for the period of 2070 to 2100 in response to just climate change—i.e. no additional effects of other anthropogenic activities. The future forecasting shows approximately 10% and 15% increases in mean EC for representative concentration pathways 2.6 and 8.5 (RCP2.6 and RCP8.5) scenarios, respectively.

This article is part of the theme issue 'Salt in freshwaters: causes, ecological consequences and future prospects'.

1. Introduction

Two types of processes can be distinguished that govern an increase of salinity: primary and secondary salinization. Primary salinization is associated with increasing salt input originating from natural processes such as rainfall, rock weathering, sea-water intrusion and aerosol deposits [1]. Human-driven salinization is called secondary salinization and is mainly induced by land development, agriculture, discharge of industrial liquid or solid waste, mining, road de-icing or intensive fertilization, and irrigation [2–4]. The effects of stream salinization on water bodies range from physiological responses of organisms [5] to alterations in freshwater communities, to subsequent reductions in ecosystem functioning [6]. Reductions in the density and species richness of multiple organism groups such as diatoms [7], macroinvertebrates [5,8], amphibians [9], fish [10] and riparian plant communities [11,12] have been reported in response to increasing surface water salinity.

Climate change can accelerate natural salinization processes, thereby increasing the salt load to water bodies. For example, modelling of climatic forcing of the water and chloride balance for freshwater lake Ijsselmeer in the Netherlands shows that the peak chloride concentrations can increase by up to 108 mg l^{-1} (an increase of 14.3% compared to the reference situation) through the climate change scenario W+ (characterized by a strong increase in the global mean temperature) in 2050. The main driver for the increase is the changing hydrology of the Rhine River expected to change owing to snowmelt and rainfall that dominate river water volume [13]; the case studies in six lakes and reservoirs in southern Europe (Estonia, Greece, Turkey and Italy) and the Middle East (Israel), and one in Brazil reveal that changes in water level owing to climate warming have significant effects on salinity level and trophic structure of lakes and reservoirs [14]. Future climate change projections for Central and eastern Europe forecast an increase in temperature extremes, such as an increase in the duration and intensity of droughts [15]. This may exacerbate salinization in this region. Although salinization in Central European countries such as Germany is currently considered rather a localized problem, mostly originating from mining [16], other studies indicated that salinization can interfere with other stressors. For example, increasing salinity was a major factor controlling the invasion of alien species (molluscs and crustaceans) in German streams [17]. Finally, other human activities may interact with and exacerbate the natural processes of salinization. Natural salts in water bodies could be enhanced through human-accelerated weathering [18–20]. For example, the disturbance of lithology through urban construction can bring bedrock materials to the surface that are subject to chemical weathering [21].

Natural background ion concentrations are driven by salt input originating from natural processes such as rainfall, rock weathering, sea-water intrusion and aerosol deposits without the presence of human activities. Models predicting baseline salinity help to establish benchmark conditions that can be used to assess whether stream water quality has degraded through secondary salinization. Moreover, such models inform on the changes and the range of variation that are likely to occur compared to baseline salinity under different scenarios of future climate change. Thereby, they also represent a first step towards investigating future salinity, including human drivers. Generally, such information is required to assess and communicate the economic costs of ongoing and future river salinization and thus to make decisions for management regarding mitigation and local adaptation.

The major factors that control the natural background level of salinity are lithology, climate, vegetation, relief and soil properties [22–27], hydrochemical processes, size and elevation of the watersheds [26–28] and groundwater [29,30]. Here, we modelled and forecast the change in natural background ion concentrations in running waters of Germany based on these major factors, focusing on electrical conductivity (EC, a proxy for total dissolved solids), Ca^{2+}, Mg^{2+} and SO_4^{2-}. Major drivers of salinization (temperature and with its resulting evaporation, dilution, etc.) are likely to change. Hence, our study evaluates the importance of the different drivers and gives a first indication of likely consequences of climate change on salinization.

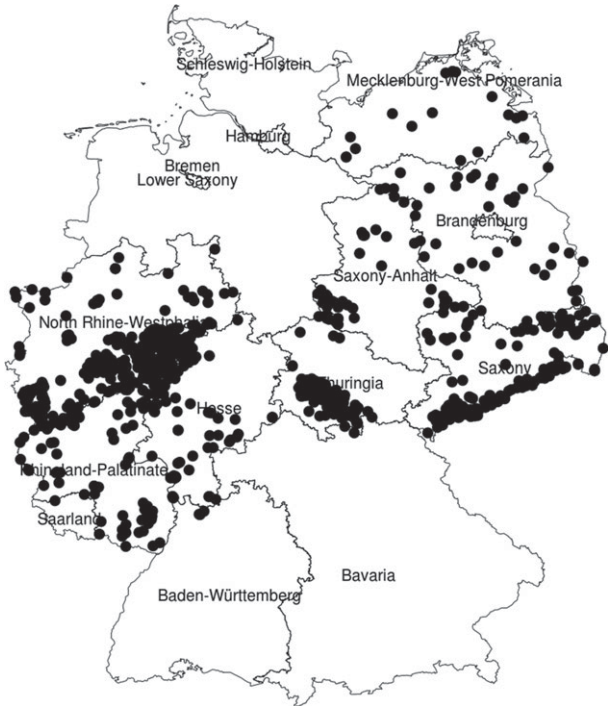

Figure 1. Location of EC monitoring sites (other ions, see electronic supplementary material, figure S1).

2. Material and methods

(a) General approach

We used the major natural factors controlling salinity such as geologic and soil properties, climate, vegetation and topography as predictors in models. We then compared the predictive power of random forest (RF) and LR models for EC. We used RF models because they can model nonlinear relationships, are insensitive to over fitting and generally have high predictive performance compared with other machine-learning methods [31–33]. However, since machine-learning approaches are often complex to interpret, we also fitted a statistical model. Since the ordinary linear regression (LR) model can perform poorly in both prediction and interpretability of results in situations with intercorrelated predictors [34], we used a penalized approach, i.e. the elastic net to overcome these drawbacks. Assessment of model performance, caveats and limitations of models are presented in subsequent sections. The model with the best predictive performance was used to forecast EC in German streams for the period from 2070 to 2100.

(b) Datasets and catchment selection

Data for EC and ion concentrations in streams, measured from 2005 to 2015, and the location of sampling sites were provided by the German federal state authorities (figure 1 and electronic supplementary material, figure S1). To identify sites that represent background EC and ion concentrations without major human influence, we selected sites according to the following two criteria: (1) less than 5% agricultural/urban land use in the catchments, (2) no mining in the catchments. These criteria of site selection were based on Olson & Hawkins [27], Herlihy *et al*. [35] and Herlihy & Sifneos [36].

The source of geologic data used in the study is The Geological Survey Map of the Federal Republic of Germany 1 : 200 000 (GÜK200) (see https://www.bgr.bund.de/EN/Themen/SammlungenGrundlagen/GG_geol_Info/Karten/Deutschland/GUEK200/guek200_inhalt_en.html, accessed on 16 July 2018). The map consists of 55 layers that give the distribution of more than 3800 geological

Table 1. Response and predictor variables. Time periods for the mean variables are described in electronic supplementary material, table S2.

response/predictor variable	category of variable	variable	unit
response	EC	electrical conductivity (EC)	mS cm^{-1}
	ion concentration	Ca^{2+}	mg l^{-1}
		Mg^{2+}	mg l^{-1}
		SO_4^{2-}	mg l^{-1}
predictor	geology	catchment mean CaO	%
		catchment mean MgO	%
		catchment mean S	%
		catchment mean unconfined compressive strength	MPa
		catchment mean log geometric mean hydraulic conductivity	10^{-6} m s^{-1}
	climate	catchment mean annual temperature	°C
		catchment mean annual precipitation	mm
		catchment mean number of freeze days	days
	soil	catchment mean available water capacity	dimensionless
		catchment mean bulk density	g cm^{-3}
		catchment mean organic matter content	%
		catchment mean soil erodibility	dimensionless
		catchment mean soil permeability (k_f)	m s^{-1}
		catchment mean soil depth	m
		catchment mean water table depth	m
	topography	catchment area	km^2
		catchment mean elevation	m
	vegetation	catchment mean enhanced vegetation index (EVI)	dimensionless
	groundwater	catchment mean recharge speed	mm $year^{-1}$

units of the surface geology of Germany and adjacent areas. The geological units contain information on the stratigraphy (age), genesis and component lithologies of the rocks.

The spatially dominant lithology was estimated for each geological unit based on 102 different lithologies listed by the GÜK200. Then, we characterized five attributes for each unit comprising percentages of CaO, MgO and S; uniaxial compressive strength as a proxy for rock strength and hydraulic conductivity as a proxy for rock–water interaction. The values of the five attributes for each lithology were derived from Olson & Hawkins [27] (electronic supplementary material, table S1).

Climatic data such as mean annual temperature, precipitation and the number of days of freeze were obtained from the DWD Climate Data Center in Germany (see https://www.dwd.de/EN/climate_environment/cdc/cdc_node.html, accessed 16 July 2018). Multi-annual grids of precipitation, air temperature (2 m above ground) and freeze days over Germany for 1981–2015 at a resolution of 1 km × 1 km were used. For further variables related to soil properties, vegetation and groundwater recharge velocity, see electronic supplementary material, table S2.

Geographical information on mining, agricultural, conservation and urban land use was extracted from the Authoritative Topographic-Cartographic Information System for Germany [37]. For each sampling site, upstream catchments were derived from a digital elevation model [26] based on the multiple flow direction algorithm [38] as implemented in GRASS GIS 7 [39,40]. During the derivation of upstream catchments, we also calculated topographic indices such as the area and elevation for each catchment. In some flat areas, the algorithm failed to delineate catchments (16% of all sites). In these cases, the catchments were assigned based on information of drainage basins provided by the Federal Institute of Hydrology for the respective stream segments, and the derived information was amalgamated with the other data [41].

(c) Modelling

There was no evidence of spatial autocorrelation, as indicated by semivariograms calculated using the R packages SSN [42] and openSTARS [43] (electronic supplementary material figure S2); therefore, no adjustments for spatial autocorrelation were needed. LR and RF models were used to develop predictive models of natural background EC and ion concentrations in streams in Germany. We compared the predictive power of both models. We considered four responses (EC, Ca^{2+}, Mg^{2+} and SO_4^{2-}) and used 19 candidate predictive variables: 5 describing geological characteristics, 3 describing climate, 7 describing soil properties, 2 capturing topography, 1 each for vegetation and groundwater (table 1).

(i) Linear regression model

The LR model is given by: $y = \beta_0 + \beta_1 \times_1 + \beta_2 \times_2 + \ldots + \beta_p x_p$, where $x_1, \ldots, x_p$ are predictors; y is the response and p is number of predictors. The vector of regression coefficients β (β_0, $\beta_2, \ldots \beta_p$) is derived in model fitting, for example, using ordinary least squares (OLS) by minimizing the residual sum of squares. However, OLS can perform poorly in both prediction and interpretability of results, particularly with intercorrelated predictors [34]. Penalized approaches such as the LASSO, ridge

regression and the elastic net have been suggested to improve OLS. The LASSO imposes an L_1-penalty on the regression coefficients and simultaneously does both continuous shrinkage and automatic variable selection such that only the important predictor variables remain in the model. On the other hand, by bounding on L_2-norm of the coefficients and continuous shrinkage, ridge regression [44] can minimize root-mean-squared errors (RMSE) and achieve higher prediction performance. The regression coefficients in these techniques are shrunk towards zero by imposing a penalty on their size [33]. Bühlmann & van de Geer [45] showed that when predictor variables are correlated, analyses with the elastic net can result in a lower mean squared error (MSE) than LASSO and ridge regression. Moreover, the use of the elastic net method has been shown to identify a higher number of influential variables than LASSO and ridge regression approaches [46].

Given that many predictors were highly correlated (electronic supplementary material, table S3), we fitted LR models by applying the elastic net (R package glmnet) [47]. The elastic net represents a combination of ridge regression and LASSO as suggested by Zou & Hastie [34] employs the elastic net penalty $P(\beta)$ composed of two component penalty functions:

$$P(\beta) = \sum_{j=1}^{p} \left(\frac{1}{2}(1-\alpha)\beta_j^2 + \alpha|\beta_j| \right). \tag{2.1}$$

The first penalty is the ridge penalty (L_2) that minimizes the weighted sum of squared regression coefficients, whereas the second component is the LASSO penalty (L_1) minimizing the weighted sum of absolute regression coefficients. The penalty parameter $\alpha \in (0,1)$ determines the bias variance trade-off between L_1 and L_2 (i.e. how much weight should be given to either the LASSO or ridge regression). The elastic net with $\alpha = 0$ performs ridge regression, whereas $\alpha = 1$ is equivalent to the LASSO; β denotes the values of the regression coefficients.

We tuned α and λ in our models and selected the optimal model as the α and λ combination (electronic supplementary material, table S4) that yielded the highest prediction performance based on five-fold cross-validation. The λ is the shrinkage parameter selected from a range of 0.0001 to 1 and hyperparameter α ranged from 0 to 1. Before developing the LR models, we applied spread-level plots (car package; [48]) to assess the residuals for heteroscedasticity which suggested that the response variable be log-transformed. We also log-transformed the predictor variables catchment area and organic matter content to improve linearity between these predictors and the response variables. Predictor variables with a larger coefficient are considered to be most important.

(ii) Random forests

RF is a machine-learning method introduced by Breiman [31] to enhance predictive accuracy and classification accuracy without overfitting data. RF is based on the principle of classification and regression trees (CART). RF uses several bootstrap samples of the data that are randomly selected at each node as a subset of explanatory variables to build many binary decision trees [31]. We used the R package randomForest for fitting RF [32].

We checked for variables that exhibit very low variance and removed highly correlated variables that were unlikely to be informative. We also used partial dependence plots to identify predictors with uninterpretable relationships with the responses, and we removed these predictors from the final models. By implementing the function tuneRF with 1500 trees, the optimal number of terminal nodes (mtry) at each node that produced the minimum out of bag error (OOB error) was determined (electronic supplementary material, table S5). In detail, a bootstrap sample of the original data was used to construct each tree in the tuneRF function. The observations that are not used to construct a tree are denoted out-of-bag (OOB) observations. OOB observations can then be predicted from the trees to evaluate prediction accuracy, where the resulting error is referred to as OOB error. The best performing model (optimal mtry) is identified as the one with minimum OOB error. Variable importance was evaluated as the mean decrease in accuracy (%IncMSE), a measure of how much the model error increases if that variable's information is removed by randomizing it.

(d) Assessment of model performance

In elastic net regression, we applied a five-fold cross-validation approach to estimate the model with the highest predictive accuracy, whereas in RF the OOB error was used to identify the best model. To compare the models, we also calculated the coefficient of determination (R^2) and the MSE of the variance for both models. The R^2 is a goodness of fit measure, whereas MSE is an absolute measure of predictive accuracy. All calculations and graphics were performed in R version 3.3.1 [49].

(e) Forecasting future EC

We examined the effects of likely climate change (temperature and precipitation) on future EC, while holding all other factors constant (i.e. geology, soil properties, vegetation and groundwater). We selected 610 standard water monitoring sampling sites in small German streams that are spatially evenly distributed and therefore representative for the whole of Germany (UBA—German Federal Environmental Protection Agency). The dataset of environmental factors for these 610 sites was extracted from the same sources as described in table 1, however, with the future data for temperature and precipitation. We applied the established RF model that showed the best predictive performance to the new dataset to predict current and forecast future EC. Then, we calculated EC alteration as the difference between the predicted 2070–2100 EC (both scenarios RCP2.6 and RCP8.5) and the baseline scenario to identify any tendencies of natural salinity to change in Germany's streams in the future.

For climate change projections, we used statistically downscaled datasets from the Delta Method for the Fifth Assessment Report of Intergovernmental Panel on Climate Change [50]. The CSIRO MK. 3.6.0 model was applied with a 30 s resolution for the time period 2070–2100 under the RCP 2.6 and RCP 8.5 (RCP—representative concentration pathways) for future climate change. While the RCP 2.6 represents one of the scenarios that aims to limit the increase of global mean temperature to 2°C, RCP 8.5 is based on a comparatively high greenhouse gas emissions scenario where the range of temperature increase is expected to be 3.5–4.5°C by 2100 [50].

3. Results and discussion

(a) Model fit, model evaluation and important variable

RF generally resulted in more parsimonious models compared with LR. The number of selected predictors in the models ranged from 12 for Mg^{2+} to 19 for both Ca^{2+} and SO_4^{2-} in the LR models, and from 7 for Ca^{2+} to 10 for SO_4^{2-} in the RF models (electronic supplementary material, table S6). We found clear differences among ions in the relative importance of the factors for predicting stream chemistry. Nevertheless, the most dominant factors in predictive models were rock chemistry and climate (electronic supplementary material, table S6). Previously, Nédeltcheva *et al*. [24] found annual rainfall and the proportion of different minerals in the bedrock to be the two main factors driving stream water chemistry, Olson & Hawkins [27,51] showed

that both rock chemistry and climate were important predictors of stream chemistry. Stream chemical composition in Queensland is a consequence of both geological history, and past and present climates [52].

Stream Ca^{2+}, Mg^{2+} and SO_4^{2-} concentrations were positively correlated with percentages of CaO, MgO and S in the catchment rock, representing the effect of different bed rock types on stream chemistry, caused primarily by the geochemical weathering and erosion. During chemical weathering, these elements become separated from the rocks as dissolved ions or colloids. Subsequently, they may be incorporated into secondary minerals or remain as primary resisted minerals. Solutions that have reacted with rocks and contain dissolved ions, colloids and suspended matter will eventually reach surface water bodies and thus increase concentrations of Ca^{2+}, Mg^{2+} and SO_4^{2-} that make up a large portion of EC. Based on catchment studies in the USA, Walling [53] found that ion concentrations in surface water were independent of solubilities of the minerals present in these rocks. Since ultrabasic rocks are rich with pyroxenes and olivine, the predominant ion expected from these rocks in surface water is Mg^{2+}. Similarly, Ca^{2+} is the dominant cationic contribution in water bodies from calcareous or dolomitic soils.

EC and ion concentrations in streams correlated negatively with mean annual precipitation. This result likely indicates that increasing precipitation results in large water volumes in streams, which causes dilution of most solutes [2,22]. Moreover, the amount of precipitation also influences the amount of flow through the soil as well as the soil water retention time before it enters streams and lakes. We found a positive relation between temperature and EC, and between temperature and ion concentrations, which may also indicate an effect of water volume. An increase in water temperature increases evaporation [54], which in turn reduces dilution capacity, translating into higher salinity [55].

Between 70% and 75% of the variance in stream water EC, Ca^{2+}, Mg^{2+} and SO_4^{2-} was explained by environmental factors in LR (figure 2 and electronic supplementary material, figure S3), respectively, with MSEs all less than 0.06. The predictive power of RF and LR models were likely similar, with RF explaining 71–76% of the spatial variation in ion concentrations and LR explaining 70–75% of the variance (figure 2 and electronic supplementary material, figure S4). The SO_4^{2-} concentration was best explained in both models ($R^2 = 76\%$, MSE = 0.033 in RF; $R^2 = 75\%$, MSE = 0.043% in LR).

(b) Forecast future EC

For Germany, climate change is expected to result in increased mean temperature and a decrease in mean precipitation until end of the century, based on the selected scenarios (electronic supplementary material, figure S5 and table S7). Evaporation is expected to rise when temperature increases [54]. A decrease in the amount of precipitation results in a decrease in catchment runoff [56]. Reduced river discharge implies a lower dilution capacity that also contributes to higher salinity concentration [55].

We forecast an approximately 10–15% increase in mean EC for the climate change scenarios for German small streams. The average ECs in the period from 2070 to 2100 for scenarios RCP 2.6 and RCP 8.5 were predicted to be 0.407 ($\pm$0.008 (= 2 s.e.)) and 0.418 ($\pm$0.007) mS cm^{-1}

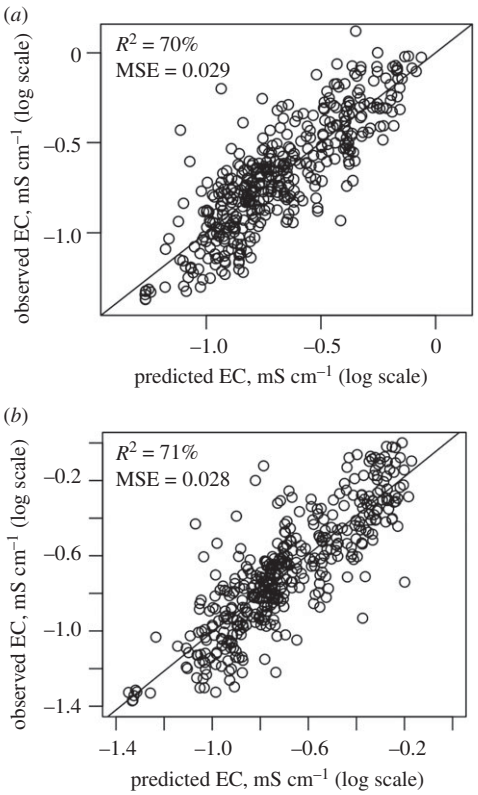

Figure 2. Plots of predicted versus observed EC value in (*a*) LR, (*b*) RF (other ions, see electronic supplementary material, figures S3 and S4).

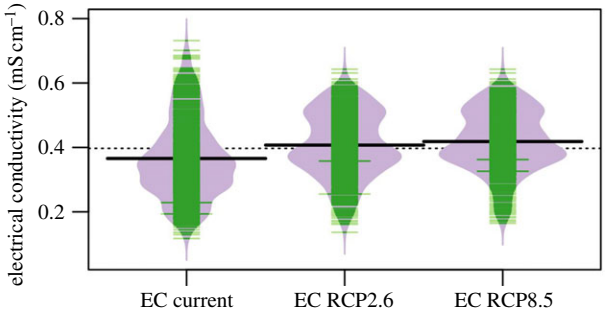

Figure 3. Bean plot of EC current values (2005–2015) and forecast values (2070–2100) in RCP 2.6 and RCP 8.5 scenarios. Thin lines give individual data points, with the length of the line indicating the relative frequency; the thick black lines gives the mean values per group; the dashed black line gives the mean value over all groups. (Online version in colour.)

compared with 0.366 ($\pm$0.010) mS cm^{-1} for the period from 2005 to 2015 (figure 3). Furthermore, the magnitude of the difference between the two scenarios in forecast EC was not as large as expected (0.407 versus 0.418 mS cm^{-1}) because precipitation was considered the most important factor in stream chemistry in the RF model, but there was only a slight difference in predicted precipitation between the two scenarios (electronic supplementary material, figure S5b).

An increase in EC was also found over the past decades in the United States (US) and Australia but mainly driven by anthropogenic factors including mining, resources extraction, agriculture and urbanization. A large proportion of the streams (37%) in the US have been impacted by increasing EC over the past 50 years [18] and the predicted rate of increase in EC is 50% [51]. In Australia, the predicted rates of average river salinity change for the period from 1998 to

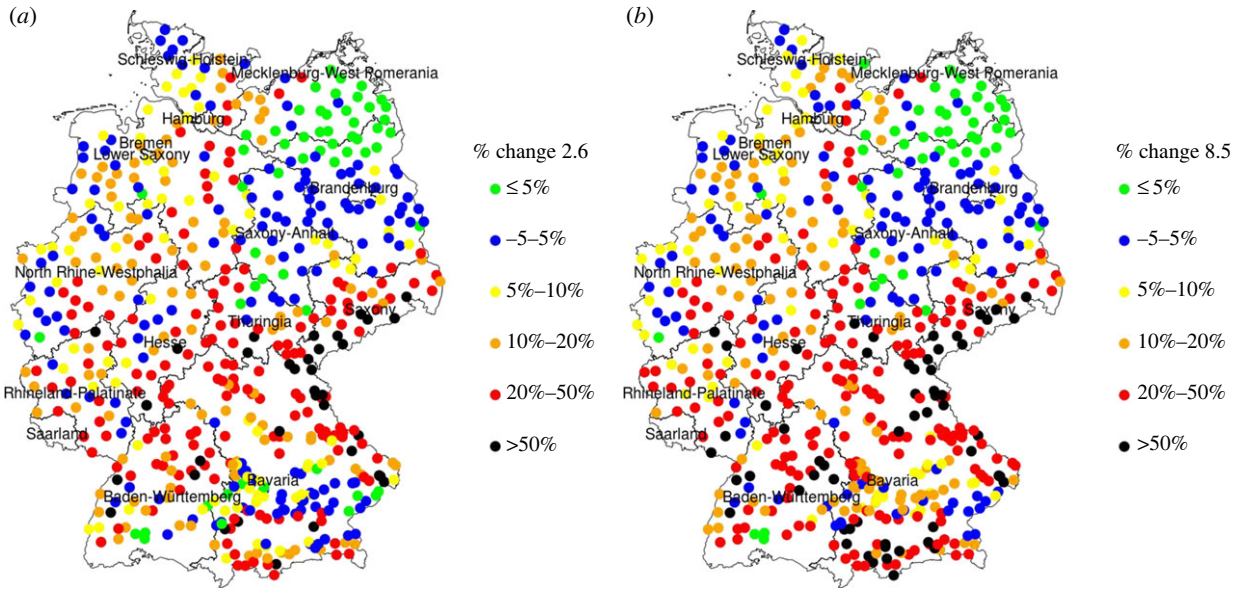

Figure 4. Percent change in mean EC (*a*) RCP 2.6 (*b*) RCP 8.5.

2100 showed strong regional differences. The average of salinity at the lower River Murray (Southeast Australia) was predicted to increase from 0.57 to 0.90 mS cm^{-1} in 100 years (i.e. 58%), whereas this rate was 60% and 505% for Central East and Northeast Australia, respectively [57,58].

Statistically increasing trends in EC were observed at approximately 80% of sites (610 sites, electronic supplementary material, figure S6) for RCP2.6 and RCP8.5, whereas this proportion is 37% for the US [18]. This finding suggests a more homogeneous response in Germany, which is not surprising given the wider gradients in climate and lithology in the US. In some cases, the predicted change in EC was substantial: for example, EC in RCP2.6 and RCP 8.5 was predicted to increase by 50% at 5% and 10% of sites, respectively, mostly in south Germany (figure 4). An increased EC can negatively impact ecosystem processes [13,55]. Populations of invertebrates (e.g *Heterocypris* sp.) may be reduced from present levels if streams salinity increases by more than 50% [59]. Shifting food habits and reducing fish production are likely consequences of a salinity-induced disruption in the benthic invertebrate forage base [59]. A large-scale study by Kefford *et al*. [5] found approximately 13%, 9%, 12%, 8%, 4% and 20% species loss in all taxa, all insects, EPT, non-EPT insects, crustaceans and molluscs, respectively, for a 30% change in EC in Southeast Australia, leading to community changes [60]. Finally, loss in community trait diversity was linear along the salinity gradient [61].

(c) Model limitations

First, owing to the selection criteria for the sampling sites (less than 5% agricultural and urban land use in the catchments, no mining) and owing to the fact that we relied on data monitored by the respective federal authorities, the sampling sites were spatially unevenly distributed and some areas of Germany lacked sampling sites (figure 1 and electronic supplementary material, figure S1). Therefore, some environments were not adequately represented in our models (especially environments in Bavaria with high elevation and more continental climates) and our predications for these areas may be biased. Second, natural salinization processes can interact with other human and non-human processes. For example, increasing oceanic alkalinity has been driven by an interplay of acidic precipitation, amongst other factors [62]. Acid rain interacts with lithology through weathering to increase dissolved inorganic carbon in river water and in turn conductivity [20]. Such factors (e.g. pH of precipitation or atmospheric deposition (Ca^{2+}, Mg^{2+}, SO_4^{2-} and other minerals)) were omitted from our model, which means that our forecast may underestimate the increase in natural background salinity. Finally, extrapolating statistical relationships outside the range of observed data should be interpreted with care. In other words, if the new predictor values are outside of the range on which the model was fit, the relationship between the independent variables and the dependent variable might change outside of that range [63]. The extreme temperatures expected in the future exceed the range of measured temperature (electronic supplementary material, figure S5) in our study, which likely causes RF to underestimate the actual change in salinity at these sites. Most climate models estimate future temperatures higher than current ones; therefore, this issue is likely to be a general one in statistical models.

4. Conclusion

We fit two statistical models (LR and RF) to forecast ion concentrations and EC in streams in Germany. The model findings may on the one hand directly inform on potential risks in other Central European regions with similar gradients in lithology and climate. On the other hand, the models may be adopted in other European regions with similar major drivers in stream chemistry to begin to establish a continent-wide assessment of both current and future changes in salinity. The results of EC projections show a slightly elevated conductivity in German streams in the period from 2070 to 2100 under climate change. Changes in other human and non-human processes such as changes in the acidity of atmospheric deposition or land use may exacerbate natural salinization processes, though incorporation of such processes was beyond the scope of this study. In particular, incorporating land use information may enhance our

predictive capacity and understanding of the future anthropogenic influence on stream salinity.

Data accessibility. The datasets supporting this article have been uploaded as part of the electronic supplementary material.

Authors' contributions. L.T.D.H.: contributions to data collection, data extraction; analysis and interpretation of data, drafting the article and revising it. M.K.: coordinated the study and helped draft the manuscript and revision. K.S.: participated in geography data analysis. J.R.O.: participated in the design of the study, helped draft the manuscript and revision. C.P.H.: helped draft the manuscript and revision. R.B.S.: substantial contributions to methodology design, helped draft the manuscript and revision. All authors gave final approval for publication.

Competing interests. We have no competing interests.

Funding. The project was funded by The Ministry of Education and Training, Vietnam.

Acknowledgement. The authors thank Eduard Szöcs for help with database compilation. The authors thank the federal state authorities, the German Federal Environmental Protection Agency for providing chemical monitoring data and the Ministry of Education and Training, Vietnam for funding for the project.

References

1. Williams W. 1987 Salinization of rivers and streams: an important environmental hazard. *J. Hum. Environ.* **16**, 180–185.
2. Williams DD, Williams NE, Cao Y. 2000 Road salt contamination of groundwater in a major metropolitan area and development of a biological index to monitor its impact. *Water Res.* **34**, 127–138. (doi:10.1016/S0043-1354(99)00129-3)
3. Ziemann H, Kies L, Schulz C-J. 2001 Desalinization of running waters: III. Changes in the structure of diatom assemblages caused by a decreasing salt load and changing ion spectra in the river Wipper (Thuringia, Germany). *Limnol. Ecol. Manag. Inland Waters* **31**, 257–280. (doi:10.1016/S0075-9511(01)80029-3)
4. Cañedo-Argüelles M, Kefford BJ, Piscart C, Prat N, Schäfer RB, Schulz C-J. 2013 Salinisation of rivers: an urgent ecological issue. *Environ. Pollut.* **173**, 157–167. (doi:10.1016/j.envpol.2012.10.011)
5. Kefford BJ, Marchant R, Schäfer RB, Metzeling L, Dunlop JE, Choy SC, Goonan P. 2011 The definition of species richness used by species sensitivity distributions approximates observed effects of salinity on stream macroinvertebrates. *Environ. Pollut.* **159**, 302–310. (doi:10.1016/j.envpol.2010.08.025)
6. Schäfer RB, Bundschuh M, Rouch DA, Szöcs E, von der Ohe PC, Pettigrove V, Schulz R, Nugegoda D, Kefford BJ. 2012 Effects of pesticide toxicity, salinity and other environmental variables on selected ecosystem functions in streams and the relevance for ecosystem services. *Sci. Total Environ.* **415**, 69–78. (doi:10.1016/j.scitotenv.2011.05.063)
7. Busse S, Jahn R, Schulz C-J. 1999 Desalinization of running waters. *Limnologica* **4**, 465–474. (doi:10.1016/S0075-9511(99)80053-X)
8. Cañedo-Argüelles M, Grantham TE, Perrée I, Rieradevall M, Céspedes-Sánchez R, Prat N. 2012 Response of stream invertebrates to short-term salinization: a mesocosm approach. *Environ. Pollut.* **166**, 144–151. (doi:10.1016/j.envpol.2012.03.027)
9. Odum WE. 1988 Comparative ecology of tidal freshwater and salt marshes. *Annu. Rev. Ecol. Syst.* **19**, 147–176.
10. Ferreri CP, Stauffer JR, Stecko TD. 2004 Evaluating impacts of mountain top removal/valley fill coal mining on stream fish populations. In *American Society of Mining and Reclamation Proceedings* (ed. R.I. Barnhisel), pp. 576–592. (doi:10.21000/JASMR04010576)
11. Lymbery AJ, Doupe RG, Pettit NE. 2003 Effects of salinisation on riparian plant communities in experimental catchments on the Collie River, Western Australia. *Aust. J. Bot.* **51**, 667–672. (doi:10.1071/BT02119)
12. Vandersande MW, Glenn EP, Walworth JL. 2001 Tolerance of five riparian plants from the lower Colorado River to salinity drought and inundation. *J. Arid Environ.* **49**, 147–159. (doi:10.1006/jare.2001.0839)
13. Bonte M, Zwolsman JJG. 2010 Climate change induced salinisation of artificial lakes in the Netherlands and consequences for drinking water production. *Water Res.* **44**, 4411–4424. (doi:10.1016/j.watres.2010.06.004)
14. Jeppesen E *et al.* 2015 Ecological impacts of global warming and water abstraction on lakes and reservoirs due to changes in water level and related changes in salinity. *Hydrobiologia* **750**, 201–227. (doi:10.1007/s10750-014-2169-x)
15. Anders I, Stagl J, Auer I, Pavlik D. 2014 Climate change in central and eastern Europe. In *Managing protected areas in Central and Eastern Europe under climate change* (eds S Rannow, M Neubert), pp. 17–30. Dordrecht, The Netherlands: Springer. (doi:10.1007/978-94-007-7960-0_2)
16. Szöcs E, Coring E, Bäthe J, Schäfer RB. 2014 Effects of anthropogenic salinization on biological traits and community composition of stream macroinvertebrates. *Sci. Total Environ.* **468–469**, 943–949. (doi:10.1016/j.scitotenv.2013.08.058)
17. Früh D, Stoll S, Haase P. 2012 Physico-chemical variables determining the invasion risk of freshwater habitats by alien mollusks and crustaceans. *Ecol. Evol.* **2**, 2843–2853. (doi:10.1002/ece3.382)
18. Kaushal SS, Likens GE, Pace ML, Utz RM, Haq S, Gorman J, Grese M. 2018 Freshwater salinization syndrome on a continental scale. *Proc. Natl Acad. Sci. USA* **115**, E574–E583. (doi:10.1073/pnas.1711234115)
19. Kaushal SS *et al.* 2017 Human-accelerated weathering increases salinization, major ions, and alkalinization in fresh water across land use. *Appl. Geochem.* **83**, 121–135. (doi:10.1016/j.apgeochem.2017.02.006)
20. Kaushal SS, Likens GE, Utz RM, Pace ML, Grese M, Yepsen M. 2013 Increased river alkalinization in the Eastern U.S. *Environ. Sci. Technol.* **47**, 10 302–10 311. (doi:10.1021/es401046s)
21. Siver PA, Canavan RW, Field CK, Marsicano LJ, Lott A-M. 1996 Historical changes in Connecticut lakes over a 55-year period. *J. Environ. Qual.* **25**, 334–345. (doi:10.2134/jeq1996.00472425002500020018x)
22. Stallard RF, Edmond JM. 1983 Geochemistry of the Amazon: 2. The influence of geology and weathering environment on the dissolved load. *J. Geophys. Res. Oceans* **88**, 9671–9688. (doi:10.1029/JC088iC14p09671)
23. Riebe CS, Kirchner JW, Finkel RC. 2003 Long-term rates of chemical weathering and physical erosion from cosmogenic nuclides and geochemical mass balance. *Geochim. Cosmochim. Acta* **67**, 4411–4427. (doi:10.1016/S0016-7037(03)00382-X)
24. Nédeltcheva TH, Piedallu C, Gégout J-C, Stussi J-M, Boudot J-P, Angeli N, Dambrine E. 2006 Influence of granite mineralogy, rainfall, vegetation and relief on stream water chemistry (Vosges Mountains, north-eastern France). *Chem. Geol.* **231**, 1–15. (doi:10.1016/j.chemgeo.2005.12.012)
25. Reimann C, Finne TE, Nordgulen Ø, Sæther OM, Arnoldussen A, Banks D. 2009 The influence of geology and land-use on inorganic stream water quality in the Oslo region, Norway. *Appl. Geochem.* **24**, 1862–1874. (doi:10.1016/j.apgeochem.2009.06.007)
26. Rothwell JJ, Dise NB, Taylor KG, Allott TEH, Scholefield P, Davies H, Neal C. 2010 A spatial and seasonal assessment of river water chemistry across North West England. *Sci. Total Environ.* **408**, 841–855. (doi:10.1016/j.scitotenv.2009.10.041)
27. Olson JR, Hawkins CP. 2012 Predicting natural base-flow stream water chemistry in the western United States. *Water Resour. Res.* **48**, W02504. (doi:10.1029/2011WR011088)
28. Pratt B, Chang H. 2012 Effects of land cover, topography, and built structure on seasonal water quality at multiple spatial scales. *J. Hazard. Mater.* **209–210**, 48–58. (doi:10.1016/j.jhazmat.2011.12.068)
29. Last WM. 1992 Chemical composition of saline and subsaline lakes of the northern Great Plains, western Canada. *Int. J. Salt Lake Res.* **1**, 47–76. (doi:10.1007/BF02904362)
30. Woocay A, Walton J. 2008 Multivariate analyses of water chemistry: surface and ground water

interactions. *Ground Water* **46**, 437–449. (doi:10.1111/j.1745-6584.2007.00404.x)

31. Breiman L. 2001 Random forests. *Mach. Learn.* **45**, 5–32. (doi:10.1023/A:1010933404324)
32. Liaw A, Wiener M. 2002 Classification and Regression by randomForest. *R News* **2**, 18–22.
33. Hastie T, Tibshirani R, Friedman J. 2009 *The elements of statistical learning: data mining, inference, and prediction*, 2nd edn. New York: NY: Springer.
34. Zou H, Hastie T. 2005 Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **67**, 301–320. (doi:10.1111/j.1467-9868.2005.00503.x)
35. Herlihy AT, Paulsen SG, Sickle JV, Stoddard JL, Hawkins CP, Yuan LL. 2008 Striving for consistency in a national assessment: the challenges of applying a reference-condition approach at a continental scale. *J. North Am. Benthol. Soc.* **27**, 860–877. (doi:10.1899/08-081.1)
36. Herlihy AT, Sifneos JC. 2008 Developing nutrient criteria and classification schemes for wadeable streams in the conterminous US. *J. North Am. Benthol. Soc.* **27**, 932–948. (doi:10.1899/08-041.1)
37. Adv. In press. Arbeitsgemeinschaft der Vermessungsverwaltungen – AdV-Online. See http://www.adv-online.de/Startseite/ (accessed on 25 May 2018).
38. Holmgren P. 1994 Multiple flow direction algorithms for runoff modelling in grid based elevation models: an empirical evaluation. *Hydrol. Process.* **8**, 327–334. (doi:10.1002/hyp.3360080405)
39. Metz M, Mitasova H, Harmon RS. 2011 Efficient extraction of drainage networks from massive, radar-based elevation models with least cost path search. *Hydrol. Earth Syst Sci* **15**, 667–678. (doi:10.5194/hess-15-667-2011)
40. Neteler M, Bowman MH, Landa M, Metz M. 2012 GRASS GIS: a multi-purpose open source GIS. *Environ. Model. Softw.* **31**, 124–130. (doi:10.1016/j.envsoft.2011.11.014)
41. WasserBLIcK. 2015 See https://wasserblick.net/servlet/is/1/ (Data downloaded 24.03.2015).
42. Hoef JV, Peterson E. 2018 *SSN: Spatial Modeling on Stream Networks*. See https://CRAN.Rproject.org/package=SSN.
43. Kattwinkel M, Szöcs E. 2018 *openSTARS: open source implementation of the STARS ArcGIS toolbox*. See https://github.com/MiKatt/openSTARS.
44. Hoerl AE, Kennard RW. 1970 Ridge regression: biased estimation for nonorthogonal problems. *Technometrics* **12**, 55–67. (doi:10.1080/00401706.1970.10488634)
45. Bühlmann P, van de Geer S. 2011 *Statistics for high-dimensional data: methods, theory and applications*. Berlin, Germany: Springer.
46. Tutz G, Ulbricht J. 2009 Penalized regression with correlation-based penalty. *Stat. Comput.* **19**, 239–253. (doi:10.1007/s11222-008-9088-5)
47. Friedman J, Hastie T, Tibshirani R, Simon N, Narasimhan B, Qian J. 2018 *glmnet: lasso and elastic-net regularized generalized linear models*. See https://CRAN.R-project.org/package=glmnet.
48. Fox J. In press. spread.level.plot function | R Documentation. See https://www.rdocumentation.org/packages/car/versions/0.8-1/topics/spread.level.plot (accessed 11 July 2018).
49. R Core Team. 2018 *R: a language and environment for statistical computing*. See https://www.gbif.org/tool/81287/r-a-language-and-environment-for-statistical-computing (accessed on 25 May 2018).
50. ICPP. 2013 IPCC Fifth Assessment Report. See https://www.ipcc.ch/report/ar5/ (accessed 15 May 2018).
51. Olson JR. 2019 Predicting combined effects of land use and climate change on river and stream salinity. *Phil. Trans. R. Soc. B* **374**, 20180005. (doi:10.1098/rstb.2018.0005)
52. McNeil VH, Cox ME, Preda M. 2005 Assessment of chemical water types and their spatial variation using multi-stage cluster analysis, Queensland, Australia. *J. Hydrol.* **310**, 181–200. (doi:10.1016/j.jhydrol.2004.12.014)
53. Walling DE. 1980 Water in the catchment ecosystem. In *Water quality in catchment ecosystems* (ed. M Gower), pp. 1–48. New York, NY: Wiley & Sons.
54. Sereda J, Bogard M, Hudson J, Helps D, Dessouki T. 2011 Climate warming and the onset of salinization: rapid changes in the limnology of two northern plains lakes. *Limnol. Ecol. Manag. Inland Waters* **41**, 1–9. (doi:10.1016/j.limno.2010.03.002)
55. Crowther RA, Hynes HBN. 1977 The effect of road deicing salt on the drift of stream benthos. *Environ. Pollut.* **14**, 113–126. (doi:10.1016/0013-9327(77)90103-3)
56. Nielsen DL, Brock MA. 2009 Modified water regime and salinity as a consequence of climate change: prospects for wetlands of Southern Australia. *Clim. Change* **95**, 523–533. (doi:10.1007/s10584-009-9564-8)
57. Murray-Darling Basin Commission (Australia), Murray-Darling Basin Ministerial Council (Australia) (eds). 1999 *The salinity audit of the Murray-Darling Basin: a 100-year perspective, 1999*. Canberra, Australia: Murray-Darling Basin Commission.
58. The Muray Darling Basin Commisson. 2015 The salinity audit of Murray-Darling Basin. See https://www.mdba.gov.au/managing-water/salinity (accessed 22 August 2018).
59. Galat DL, Coleman M, Robinson R. 1988 Experimental effects of elevated salinity on three benthic invertebrates in Pyramid Lake, Nevada. *Hydrobiologia* **158**, 133–144. (doi:10.1007/BF00026272)
60. Kefford BJ, Schäfer RB, Liess M, Goonan P, Metzeling L, Nugegoda D. 2010 A similarity-index–based method to estimate chemical concentration limits protective for ecological communities. *Environ. Toxicol. Chem.* **29**, 2123–2131. (doi:10.1002/etc.256)
61. Schäfer RB, Kefford BJ, Metzeling L, Liess M, Burgert S, Marchant R, Pettigrove V, Goonan P, Nugegoda D. 2011 A trait database of stream invertebrates for the ecological risk assessment of single and combined effects of salinity and pesticides in South-East Australia. *Sci. Total Environ.* **409**, 2055–2063. (doi:10.1016/j.scitotenv.2011.01.053)
62. Müller JD, Schneider B, Rehder G. 2016 Long-term alkalinity trends in the Baltic Sea and their implications for CO_2-induced acidification. *Limnol. Oceanogr.* **61**, 1984–2002. (doi:10.1002/lno.10349)
63. Conn PB, Johnson DS, Boveng PL. 2015 On extrapolating past the range of observed data when making statistical predictions in ecology. *PLoS ONE* **10**, e0141416. (doi:10.1371/journal.pone.0141416)