Check for updates

JGR Biogeosciences

RESEARCH ARTICLE
10.1029/2024JG008031

Key Points:
- Environmental monitoring data and hydrological model outputs can be used to quantify stream CO_2 concentrations and fluxes at national scale
- Machine learning models can be trained to predict seasonal stream CO_2 concentrations based on catchment and stream characteristics
- The most important drivers of stream CO_2 are related to landscape morphometry and soil-groundwater-stream connectivity

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

K. Sand-Jensen,
ksandjensen@bio.ku.dk

Citation:

Martinsen, K. T., Sand-Jensen, K., Bergmann, V., Skjærlund, T., Kjær, J. E., & Koch, J. (2024). Seasonal carbon dioxide concentrations and fluxes throughout Denmark's stream network. *Journal of Geophysical Research: Biogeosciences*, *129*, e2024JG008031. https://doi.org/10.1029/2024JG008031

Received 22 JAN 2024
Accepted 1 JUL 2024

Author Contributions:

Conceptualization: Kenneth Thorø Martinsen, Kaj Sand-Jensen, Julian Koch
Data curation: Kenneth Thorø Martinsen, Victor Bergmann, Tobias Skjærlund, Johan Emil Kjær, Julian Koch
Formal analysis: Kenneth Thorø Martinsen, Kaj Sand-Jensen, Julian Koch
Funding acquisition: Kaj Sand-Jensen
Investigation: Kenneth Thorø Martinsen, Victor Bergmann, Tobias Skjærlund, Johan Emil Kjær
Methodology: Kenneth Thorø Martinsen, Julian Koch

© 2024. The Author(s).
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Seasonal Carbon Dioxide Concentrations and Fluxes Throughout Denmark's Stream Network

Kenneth Thorø Martinsen[1], Kaj Sand-Jensen[1], Victor Bergmann[1], Tobias Skjærlund[1], Johan Emil Kjær[1], and Julian Koch[2]

[1]Freshwater Biological Laboratory, Department of Biology, University of Copenhagen, Copenhagen, Denmark,
[2]Department of Hydrology, Geological Survey of Denmark and Greenland, Copenhagen, Denmark

Abstract Streams are important freshwater habitats in large-scale carbon budgets because of their high CO_2 fluxes which are driven by high CO_2 concentrations and surface-water turbulence. High CO_2 concentrations are promoted by terrestrial carbon inputs, groundwater flow, and internal respiration, all of which vary greatly across space and time. We used environmental monitoring data to calculate CO_2 concentrations along with a wide range of predictor variables including outputs from a national hydrological model and trained machine learning models to predict spatially distributed seasonal CO_2 concentrations in Danish streams. We found that streams were supersaturated in dissolved CO_2 (mean = 118 μM) and higher during autumn and winter than during spring and summer. The best model, a Random Forest model, scored $R^2 = 0.46$, MAE = 46.0 μM, and $\rho = 0.72$ on a test set. The most important predictor variables were catchment slope, seasonality, height above nearest drainage, and depth to groundwater, highlighting the importance of landscape morphometry and soil-groundwater-stream connectivity. Stream CO_2 fluxes determined from the predicted concentrations and gas transfer velocities estimated using empirical relationships averaged 253 mmol m^{-2} d^{-1}, and the annual emissions were 513 Gg CO_2 from the national stream network (area = 139 km^2). Our analysis presents a framework for modeling seasonal CO_2 concentrations and estimating fluxes at a national scale by means of large-scale hydrological model outputs. Future efforts should consider further improving the temporal resolution, direct measurements of fluxes and gas transfer velocities, and seasonal variation in stream surface area.

Plain Language Summary Streams play an essential role in the global carbon cycle as they usually are very rich in CO_2 and thus emit much CO_2 to the atmosphere To estimate how much CO_2 is emitted from streams, we used environmental data to calculate CO_2 concentrations and a wide range of terrain and stream flow characteristics. We used this data to train a machine learning model to predict seasonal stream CO_2 concentrations across Denmark. We found that CO_2 concentrations in Danish streams were generally higher during autumn and winter than during spring and summer. The key factors determining CO_2 concentrations included terrain slope, seasonality, elevation relative to nearby streams, and depth to groundwater. From the national stream CO_2 concentrations, we estimated the exchange of CO_2 between stream and atmosphere and the annual emissions. This study demonstrates how machine learning models and data from multiple sources can be used to predict stream CO_2 concentrations at large scales. The framework and model allow us to quantify and manage the CO_2 emissions from streams at a national scale.

1. Introduction

Freshwater ecosystems, particularly running waters such as rivers and streams (the term streams used from here), play a crucial role in the global carbon cycle (Raymond et al., 2013). They link terrestrial habitats to the sea, and facilitate the transport and processing of carbon (Cole et al., 2007). Mineralization of organic matter within streams produces carbon dioxide (CO_2), which, along with substantial external sources, contributes to the commonly observed supersaturation of dissolved CO_2 (Hotchkiss et al., 2015). This makes streams open windows of CO_2 emission to the atmosphere (Butman et al., 2016; Wallin et al., 2013). Spatiotemporal CO_2 dynamics comprise a complex interplay of biotic and abiotic factors.

The extensive variability of CO_2 concentrations in streams are well documented (Marx et al., 2017; Sand-Jensen, Riis & Kjær et al., 2022). This variability is primarily driven by the dynamic nature of groundwater and terrestrial contributions (Duvert et al., 2018; Humborg et al., 2010), and by the hydrology and biological processes that influence both the flow of water and organic matter in streams and the gas flux across the air-water interface

Writing – original draft: Kenneth Thorø Martinsen, Kaj Sand-Jensen, Julian Koch
Writing – review & editing: Kenneth Thorø Martinsen, Kaj Sand-Jensen, Victor Bergmann, Tobias Skjærlund, Johan Emil Kjær, Julian Koch

(Hutchins et al., 2019; Long et al., 2015). These processes are subject to seasonal variations which can be attributed to direct climatic drivers, such as temperature and runoff fluctuations, or indirect effects mediated by local riparian or catchment-scale factors, like terrestrial vegetation cover and land use changes. Consequently, incorporating seasonality is crucial for enhancing the predictive capabilities of current models for CO_2 concentrations in streams. Current methods to predict spatiotemporal stream CO_2 concentrations are limited and the temporal component is often neglected (Lauerwald et al., 2015; Martinsen et al., 2020) or only cover one or few catchments. One recent exception is Liu et al. (2022), who predicted monthly CO_2 concentrations in streams at a global scale. To address this gap, our study combines traditional spatial data with hydrological model outputs to improve seasonal predictions of stream CO_2 concentration.

Hydrology influences stream CO_2 concentrations. In particular, water velocity and discharge influence CO_2 concentrations directly, because they drive water turbulence and gas transfer velocity and thus CO_2 exchange with the atmosphere. Furthermore, discharge also influences lateral inputs of organic matter and dissolved CO_2 from land (Liu & Raymond, 2018). Groundwater is generally CO_2 supersaturated and contributions are constrained by local hydrogeological and topographical settings (Crawford et al., 2014; Duvert et al., 2018). Partitioning discharge into the constituting flow components could potentially improve predictions of CO_2 concentrations because different flow components are expected to have different CO_2 concentrations (Sand-Jensen, Riis & Kjær et al., 2022). Detailed hydrological models that integrate groundwater and surface water processes are required for quantifying flow contributions in space and time. By combining traditional spatial data with hydrological model outputs, we aim to provide more robust and process-constrained seasonal predictions of stream CO_2 concentrations and gas transfer velocities, ultimately enabling better estimates of spatiotemporally upscaled CO_2 fluxes.

Machine learning algorithms offer a promising approach to predicting CO_2 concentrations in streams (Martinsen et al., 2020). By leveraging large data sets of monitoring data and relevant predictor variables, machine learning models can learn patterns and relationships that are not easily identifiable through traditional statistical methods (Breiman, 2001). The advantages include improved accuracy and flexibility in modeling nonlinear relationships, as well as their interactions (Olden et al., 2008). However, model interpretation can be challenging due to the complexity of the algorithms (James et al., 2013). Traditional approaches rely heavily on simplifying assumptions and parameterizations, whereas machine learning models are less transparent in their calculations. Ultimately, integrating machine learning with the outputs of traditional modeling frameworks could lead to more accurate and robust predictions of stream CO_2 concentrations, offering insights into the complex processes governing carbon cycling in freshwater ecosystems and permitting upscaling of fluxes to regional and national scales.

In this study, we compile country-level data consisting of static catchment characteristics and dynamic hydrological variables derived from a national hydrological model. With this data, we apply machine learning methods to predict seasonal CO_2 concentrations throughout the Danish stream network. We hypothesize that: (a) seasonality is an important predictor of CO_2 concentrations in streams, and (b) hydrological flow components, in particular groundwater inflow, influence both CO_2 delivery and emissions and should therefore improve predictability of CO_2 concentrations in streams. We use the predicted CO_2 concentrations and hydrological data to estimate CO_2 fluxes and compare these with in situ measurements. Finally, we upscale CO_2 fluxes from the entire Danish stream network.

2. Materials and Methods

We employed a model selection approach to identify the optimal model for predicting seasonal CO_2 concentrations. CO_2 concentrations were derived from observations of pH and alkalinity obtained through the national environmental monitoring program. Additionally, a suite of hydrological and landscape environmental variables, expected to influence stream CO_2 dynamics, were included as predictor variables (Table S1 in Supporting Information S1). CO_2 observations were aligned with the locations along the stream network where the national hydrological model of Denmark (DK-Model), simulates discharge and flow components. For both the monitoring and DK-Model outputs, we used data spanning a range of 10 years from 2000 to 2009. The predicted CO_2 concentrations were used to estimate the CO_2 flux using the hydrological model components and aggregated to obtain national CO_2 emission estimates. The estimated seasonal CO_2 fluxes are compared to in situ measurements.

21698961, 2024, 7, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024JG008031 by Royal Danish Library, Wiley Online Library on [07/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

21698961, 2024, 7, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024JG008031 by Royal Danish Library, Wiley Online Library on [07/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.1. Study Region

The topography of Denmark is generally low-relief with elevations above sea-level ranging from −18 to 172 m, mean annual precipitation of 2.3 mm d^{-1}, and mean air temperature of 9.0°C, during the study period. The majority of soils are calcareous in nutrient-rich moraine landscapes, except for SW-Jutland which was not covered during the Weichselian glaciation and has mainly well-leached, sandy soils lower in carbonate and clay minerals (Figure 1a). Due to mineral weathering, this has resulted in differences in water chemistry between the less alkaline SW-Jutland and more alkaline major parts of Denmark. The land use is dominated by agriculture (approx. 60%).

2.2. Predicting Stream CO_2 Concentrations

2.2.1. CO_2 Concentrations

Stream CO_2 concentrations were calculated from daytime measurements of alkalinity, pH, and water temperature sampled as part of the national environmental monitoring program (Lauridsen et al., 2005). Specifically, we calculated CO_2 concentrations using the AquaEnv R-package (Hofmann et al., 2010). Estimates of CO_2 from alkalinity and pH are subject to high uncertainty in low-alkalinity (<1 meq L^{-1}), acidic, often humic waters, with an increasing degree of overestimation as alkalinity decreases further toward zero (Abril et al., 2015). However, Danish catchments are generally rich in carbonate minerals resulting in alkaline waters, and show close agreement between measured CO_2 concentrations and those estimated from alkalinity and pH (Sand-Jensen, Riis & Kjær et al., 2022). In our data, 21.6% of observations were below 1 meq L^{-1} and 4% below 0.5 meq L^{-1}. To avoid unrealistically high values, we excluded observations with low pH (<5.4; 0.6% of observations) and very high estimated CO_2 concentrations (40,000 µatm; 0.05% of observations) similar to other studies (Hastie et al., 2018; Martinsen et al., 2020). We determined seasonal means for sites with four or more observations resulting in 745 seasonal CO_2 concentrations from 309 sites for further analysis.

2.2.2. Catchment Delineation

We delineated the topographical catchment areas for each of the 62,726 DK-Model surface water simulation points (Q-points) using a digital elevation model (DEM) with a resolution of 10 m, resampled from a high-resolution (1.6 m) national DEM (SDFI, 2021). DEM processing and catchment delineation were carried out using the WhiteboxTools and TauDEM software (Lindsay, 2016a, 2016b; Tarboton, 2017). Specifically, the DEM was preprocessed to remove hydrological sinks, areas where water would accumulate due to cells being surrounded by higher elevations, using a hybrid breaching and filling approach (Lindsay, 2016a, 2016b). This was followed by determination of flow directions using the deterministic-eight flow scheme (O'Callaghan & Mark, 1984). We extracted a virtual stream network and snapped DK-model Q-points to this network with a maximum distance of 100 m and delineated the catchment area.

2.2.3. Dynamic Hydrological and Meteorological Variables

We included hydrological and meteorological variables for each stream site and season. Simulated discharge and four flow components from the DK-Model were used for further analysis. The DK-Model is a physically based and spatially distributed hydrological model that, using the MIKE SHE model code, integrates groundwater, surface water, and anthropogenic activities. The DK-Model has been developed by the Geological Survey of Denmark and Greenland over the last 25 years (Henriksen et al., 2023; Højberg et al., 2013; Soltani et al., 2021) and is currently used in basic research and a range of applications such as national water resources assessments, status of water resources, and adaptation to climate change. The DK-model simulates daily flow components and total discharge at 62,726 Q-points throughout the stream network. The flow components are simulated locally and are added to the total discharge which aggregates all upstream flow contributions. The components are (a) overland inflow (OF), (b) groundwater inflow (GWF), (c) drainage inflow from groundwater (DF), and (d) and drainage inflow from ponded water (DPF) (Figure S1 in Supporting Information S1). OF represents water flow on the terrain following topographical gradients and the process is initiated in a grid cell when the overland storage exceeds predefined detention storage. GWF represents the direct exchange between groundwater and surface water. Most streams in Denmark are associated with a positive GWF, indicating that groundwater contributes as baseflow to the discharge generation. Such a condition is referred to as gaining streams and initiated by an upward groundwater gradient in the stream channel. DF represents water that moves from groundwater storage to local

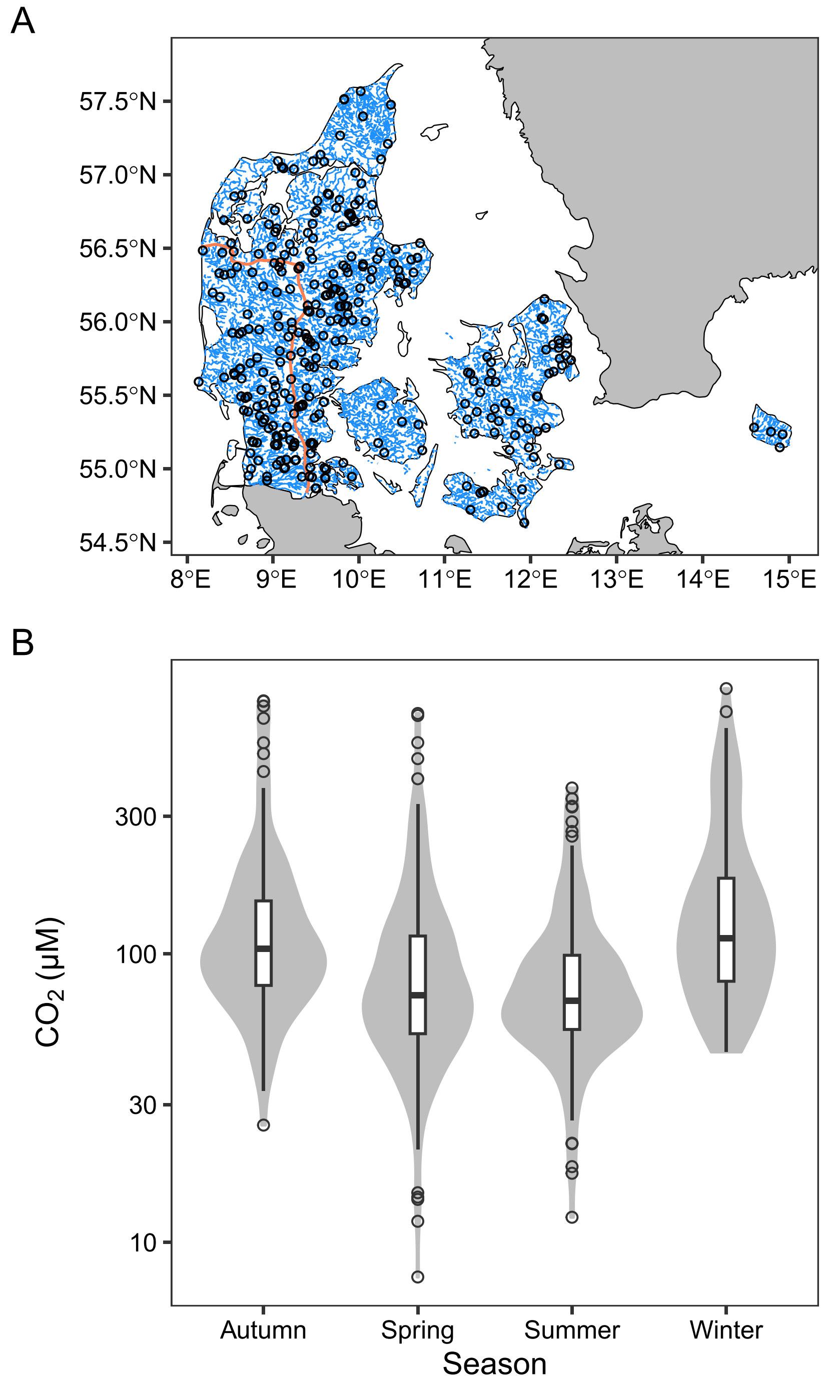

Figure 1. (a) Map showing the stream network included in the DK-model (blue) and stream sites (open points) with one or more observations of seasonal CO_2 concentration. The orange line represents the maximum extent of the ice sheet during the most recent ice age where eastern Denmark was covered by ice and SW-Jutland remained ice-free. (b) Density distributions (gray) and boxplots of seasonal CO_2 concentrations. Box plots show median (solid horizontal line), 25% and 75% quartiles (upper and lower hinge), lines extending to one and a half times the inter-quartile range (upper and lower whisker), and observations outside this range (points).

21698961, 2024, 7, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024JG008031 by Royal Danish Library, Wiley Online Library on [07/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

21698961, 2024, 7, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024JG008031 by Royal Danish Library, Wiley Online Library on [07/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

surface waters and DF is initiated in a grid cell when the groundwater level exceeds a predefined drain depth. Within the DK-model, DF is expected to represent tile drainage from agricultural fields. Since detailed data on the location and effectiveness of tile drains are absent at national scale, the drain flow process representation is calibrated against streamflow observations to obtain effective parameter values. DPF represents processes of urban drainage networks and ponded water that moves to local surface waters. Typically, DF constitutes a major part of the generated flow and is highly variable in time with the largest flows occurring during winter and spring when groundwater levels are highest. GWF varies less in time than DF and contributes much of the flow generated during summer. OF and DPF are more variable in time and space and are driven by intense precipitation events.

We removed Q-points that were located within lakes (3,172), specific to the DK-model (2,741), more than 100 m away from the virtual stream network (3,810), and/or did not intersect the DEM (613). For each DK-model Q-point, daily discharge and the four flow components were averaged seasonally. From national climate grids, we extracted seasonal mean precipitation (10 km grid) and air temperature (20 km grid) for each catchment (Scharling, 1999a, 1999b).

2.2.4. Static Variables

The static attributes are stated in three different ways: (a) as average values of relevant map layers over the catchment area, (b) as site-specific values, or 3) as locally aggregated statistics using a 200 m buffer analysis. These map layers included for example, slope and hydrological landscape indices such as height above the nearest drainage (HAND), derived from the 10 m DEM. HAND represents the vertical distance between a terrain grid cell and the stream grid cell it drains to (Rennó et al., 2008). We extracted land use from a national 10 m data set (Levin, 2022), geology-related variables (Adhikari et al., 2013), and the simulated groundwater depth averaged over time of the DK-Model. Furthermore, we identified land use in the near upstream area using the intersection between a circular 200 m buffer zone and the catchment for each site. See Table S1 in Supporting Information S1 for a detailed overview of the predictor variables.

2.2.5. Modeling

We split the data into a training (80%) and test (20%) data set. We grouped observations by stream site such that observations from the same site were not split during resampling. Inter-correlation between numeric predictor variables was assessed using Pearson correlation. Before modeling, the “season” predictor variable was one-hot encoded, and numeric predictor variables were preprocessed by applying the Yeo-Johnson power transformation followed by zero-mean and unit variance standardization. Using the training data, we explored the performance of several different models by 5-fold cross-validation (outer loop) to select the best model. Many machine learning models require tuning of hyperparameters for optimal performance. We defined hyperparameter search spaces which were sampled using 50 iterations of random search and evaluated using 5-fold cross-validation (inner loop). Model selection was performed using nested cross-validation. The best model was then trained on the entire training data set, evaluated on the test set, and used to make predictions for all DK-model sites (53,668). To assess model performance, we determine the R^2, root mean squared error (RMSE), mean absolute error (MAE), Pearson correlation coefficient (ρ), and mean absolute percentage error (MAPE). The influence of predictor variables on CO_2 concentrations in the best model was investigated using permutation variable importance and partial dependence plots. To explore the influence of the number of training observations on model performance we assessed cross-validated performance using random subsets of the training data of different sizes. Machine learning models were trained and evaluated using the Python scikit-learn library (Pedregosa et al., 2011).

2.3. Estimating Stream CO_2 fluxes

2.3.1. CO_2 flux Estimates

The CO_2 flux is the product of the gas transfer velocity (k) and the air-water concentration gradient:

$$F = \mathrm{k}(\mathrm{CO}_{2-\mathrm{water}} - \mathrm{CO}_{2-\mathrm{sat}}) \tag{1}$$

We estimated CO_2 flux from the predicted seasonal CO_2 concentrations and empirical relationships for k assuming an atmospheric partial pressure of 400 μatm for the study period 2000–2009. Partial pressure was converted to concentration using Henry's Law as a function of water temperature. Seasonal water temperature was estimated for all sites using a fitted linear relationship between observed water temperatures and air temperatures (Figure S2 in Supporting Information S1). k_{600} (m d^{-1}), k normalized to a Schmidt number of 600 (CO_2 at 20°C), was estimated as a function of water velocity (v, m s^{-1}) and slope (s, dimensionless) based on the relationship ($k_{600} = v \times s \times 2{,}841 + 2.02$) from Raymond et al. (2012). k at the ambient temperature is determined using the ratio of Schmidt numbers based on relationships in Jähne et al. (1987) and Wanninkhof (1992). Water velocity was estimated from DK-model derived discharge (m^3 s^{-1}) using hydraulic geometrical relationships ($v = 0.19 \times Q^{0.29}$; Raymond et al., 2012), and the slope was determined for each stream segment of the virtual stream network using TauDEM. CO_2 fluxes could be estimated for 53.653 DK-model q-points.

2.3.2. CO_2 Flux Measurements

To validate the estimated CO_2 fluxes we measured daytime CO_2 fluxes using floating chambers multiple times across Denmark (Rivers Tryggevælde, Odense, Omme, and Tude) resulting in 54 measurements from 27 stream sites. Floating chambers were equipped with small, inexpensive sensors (SenseAir, Sweden) as described in Bastviken et al. (2015) measuring the headspace CO_2 partial pressure every 30 or 60 s for at least 20 min. The flux was determined by the change in CO_2 partial pressure over time:

$$F = \frac{dCO_2}{dt} \frac{VP}{RTA} \tag{2}$$

where dCO_2/dt is the slope estimate by linear regressions, P is the atmospheric pressure (atm), V is the chamber volume (0.008 m^3), A is the chamber area (0.075 m^2), R is the gas constant (m^3 atm K^{-1} mol^{-1}) and T is the temperature (Kelvin). We used the average flux determined from two or three chambers at each visit. On most occasions, we measured pH in the field (YSI Professional Plus, USA) and collected water samples for measuring alkalinity using Gran titration (Gran, 1952) and to calculate the CO_2 concentration. We also determined k from the observed CO_2 concentration and flux according to Equation 1, and normalized k to k_{600} for comparisons.

2.3.3. Upscaling CO_2 Fluxes

We determined the annual CO_2 flux from the Danish stream network using the estimated fluxes. We sampled the stream network at 100 m resolution (191.165 sites), estimated stream width for all sites, and used nearest neighbor interpolation to determine the flux. Stream width was determined as a function of catchment area using empirical relationships in Denmark for six geographical regions (Table S2 in Supporting Information S1; Olsen & Højberg, 2011). Based on this analysis, the average stream width was 7.3 m with a standard deviation of 6.5 m. Since the relationships were based on the catchment area, the stream area did not differ between seasons. Finally, the annual flux was determined by multiplying the flux, stream width, and resolution used for sampling the stream network before aggregating by season. The upscaling procedure proved insensitive to the applied sampling resolution.

3. Results

3.1. CO_2 Concentrations

For the 745 seasonal observations used in the analysis, the mean alkalinity was 2.6 (range 0.05–7.6) meq L^{-1} with a pH of 7.7 (range 6.0–8.7). The mean CO_2 concentration was 118 (range 8–832) μM with significant seasonal differences (One-way ANOVA, $F = 21.2$, p-value <0.0001). Concentrations were lower in spring and summer compared to autumn and winter but there were no pairwise differences between winter versus autumn or spring versus summer (Figure 1b). 737 observations (98.9%) had CO_2 concentrations above air saturation, highlighting the prevalence of CO_2 supersaturation in streams.

3.2. Modeling CO_2 Concentrations

We evaluated the performance of several models for predicting seasonal stream CO_2 concentrations from catchment characteristics, hydrology, and climate (Table S3 in Supporting Information S1). Many of the models performed poorly indicating that relationships between CO_2 concentrations and predictor variables were complex. Furthermore, the performance of the 5 cross-validation folds was subject to high variability. Based on all performance metrics, the Random Forest model showed superior performance and was therefore selected for

21698961, 2024, 7, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024JG008031 by Royal Danish Library, Wiley Online Library on [07/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

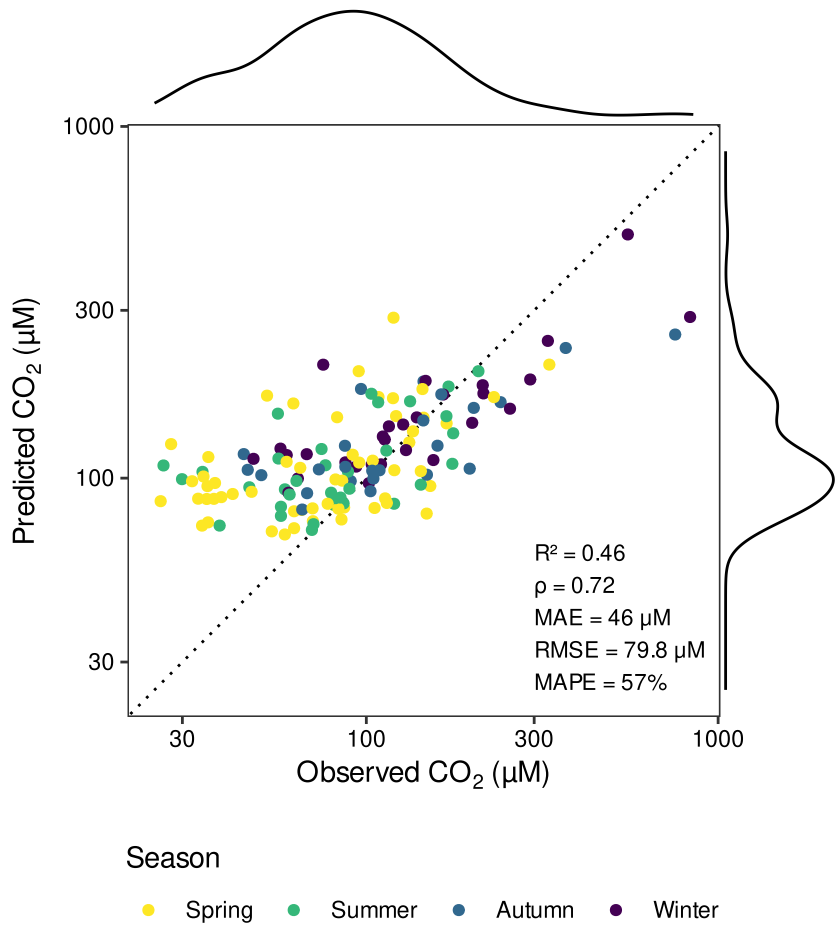

Figure 2. Observed (*x*-axis) and predicted values (*y*-axis) of stream CO_2 concentrations (μM) from the test set colored by season with density distributions on the margins. Predicted CO_2 concentrations are based on a Random Forest model.

further analysis. Following training on the entire training data, the Random Forest model was evaluated on the test set ($R^2 = 0.46$, MAE = 46.0 μM, RMSE = 79.8 μM, ρ = 0.72, and MAPE = 57%). The model appeared to over-predict at lower CO_2 concentrations which was most common in spring and summer (Figure 2).

The model was used to make seasonal predictions for all DK-model Q-points (Figure 3). At the regional scale, predicted CO_2 concentrations were related to geology with high concentrations in the glacial outwash plains west of the main stationary line of the most recent ice age, that is, SW-Jutland (Figure 1a) and in the southern islands with low relief landscape. The performance estimated during model selection (Table S3 in Supporting Information S1) and the performance determined on the test set were quite different. We investigated the sensitivity of the model to different sizes of training data and found that performance increased with the size of the training data and variability between cross-validation folds remained substantial (Figure S3 in Supporting Information S1).

We explored the importance of the predictor variables and their relationship to stream CO_2 concentrations. The most important predictor was the mean catchment slope, followed by HAND, and air temperature (Figure 4a). The functional relationships, without considering interactions, were generally as expected (Figure 4b). The influence of slope and HAND was most pronounced at low mean catchment slopes and HAND, indicating that CO_2 concentrations are higher in streams in flat terrain where contact between stream water and soils is likely to be high. The response to temperature followed the expectation that CO_2 concentrations would be greater during winter and lower in summer. Several of the predictor variables, for example, catchment slope, were related to other predictor variables (Figure S4 in Supporting Information S1).

21698961, 2024, 7, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024JG008031 by Royal Danish Library, Wiley Online Library on [07/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

21698961, 2024, 7, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024JG008031 by Royal Danish Library, Wiley Online Library on [07/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

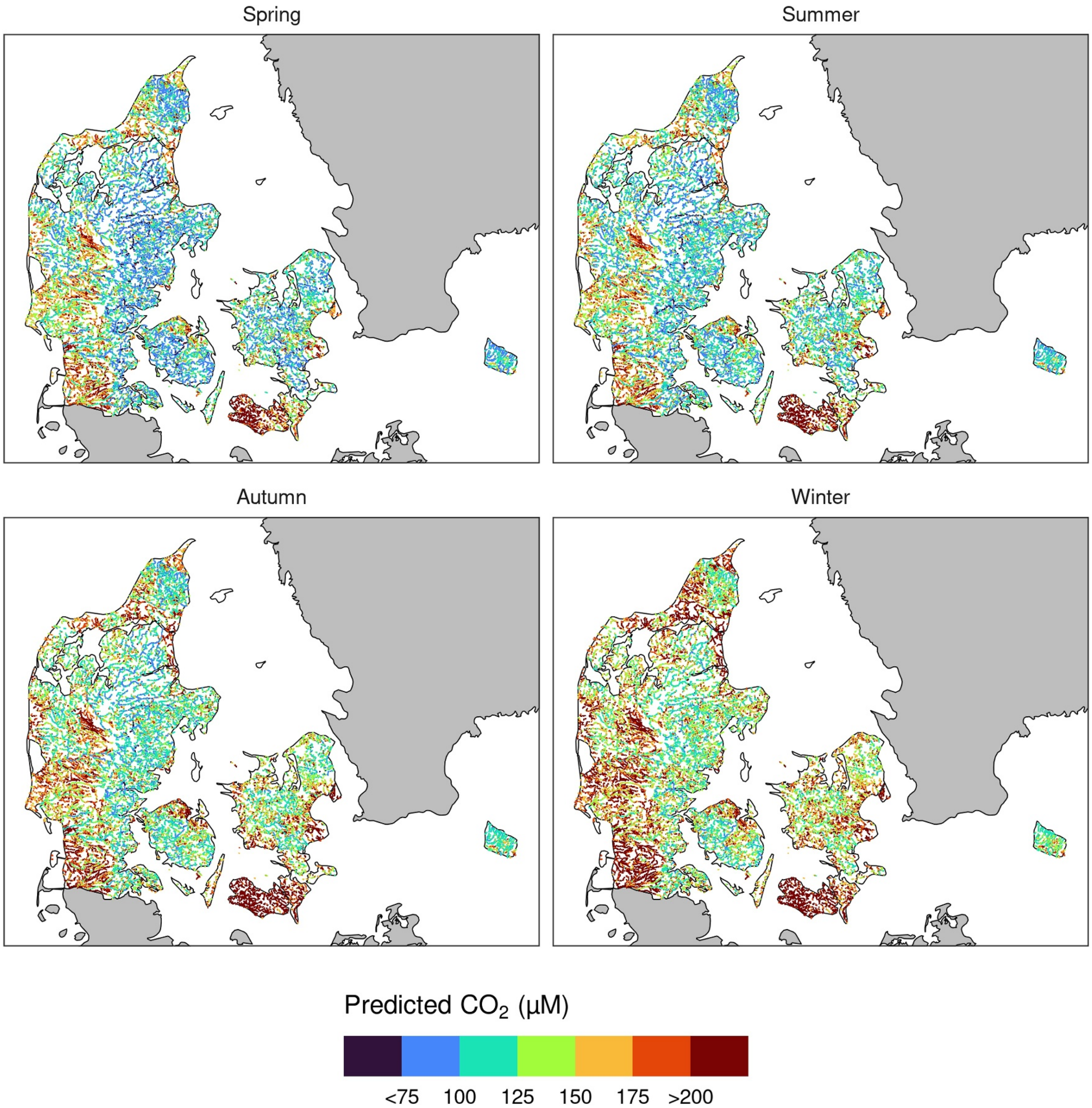

Figure 3. Predictions of CO_2 concentration for each season (a–d) covering Denmark based on a Random Forest model. The stream network consists of sites included in the hydrological DK-model.

3.3. CO_2 Fluxes

CO_2 fluxes determined from the predicted CO_2 concentration and estimated k had an overall mean of 253 mmol m^{-2} d^{-1} with the highest fluxes during winter and lowest during spring (Table 1). We estimated the CO_2 flux for all sites to produce national maps of seasonal CO_2 fluxes (Figure 5). The maps highlight regions with high CO_2 fluxes, for example, SW-Jutland and Lolland-Falster, which coincide with areas with high predicted CO_2 concentrations (Figure 3). Spatial patterns of CO_2 flux within the stream network were consistent with expectations. Lower CO_2 concentrations were observed at lake outlets during summer and increased gradually with increasing distance downstream from the outlets (Figure 6).

Floating chamber measurements of CO_2 fluxes had a mean of 186 mmol m^{-2} d^{-1} and a larger variation than the estimated flux (range 80–349 mmol m^{-2} d^{-1}). We found good agreement between observed and estimated CO_2 fluxes for some of the observations, while others had been substantially overestimated (Figure 7a). Observed and estimated CO_2 concentrations showed better agreement than the corresponding k-values.

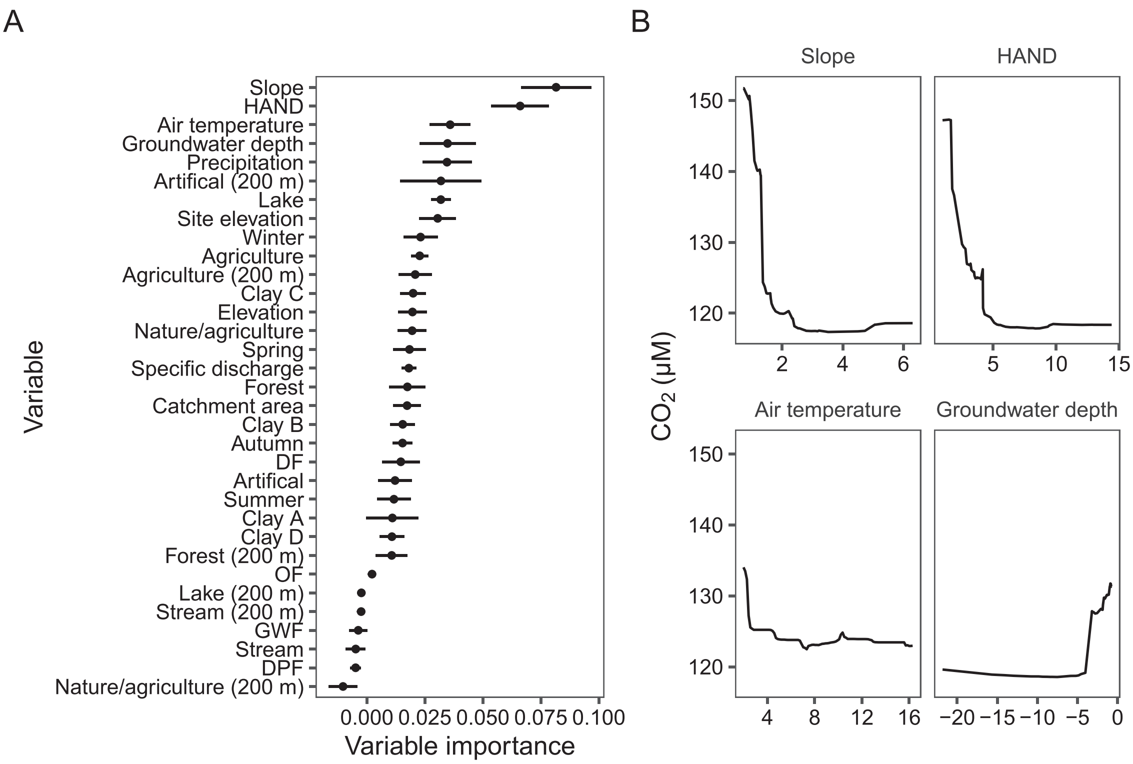

Figure 4. (a) Mean (point) ± standard deviation (line range) permutation variable importance computed on the test set for all predictor variables. (b) Relationship between predicted CO_2 concentration (y-axis) and the four most important predictor variables (x-axis) determined using partial-dependence plots.

Based on the national stream CO_2 flux estimates we calculate aggregated national estimates of 512.6 Gg CO_2 year^{-1} (spring = 440.3, summer = 513.4, autumn = 541.2, and winter = 555.2 Gg CO_2 year^{-1}). The total stream area was 139 km^2 yielding an annual average of 3,675 g CO_2 m^{-2} stream y^{-1} or 11.9 g CO_2 m^{-2} y^{-1} for Denmark's total surface area (43,100 km^2).

4. Discussion

4.1. Predicting Stream CO_2 Concentrations

We show how a machine learning approach can be used to predict stream CO_2 concentrations from catchment features, similar to other studies (Lauerwald et al., 2015; Martinsen et al., 2020) and expand with two key aspects, namely adding seasonal resolution and including dynamic hydrological features from a national scale hydrological model. The predicted CO_2 concentrations in Danish streams are as expected generally supersaturated (Rebsdorf et al., 1991) and lowest in spring and summer and higher during autumn and winter (Jones & Mulholland, 1998; Sand-Jensen & Staehr, 2012). This influence of seasonality was evident from the response to air temperature with highest CO_2 concentrations in winter at low surface irradiance and photosynthetic activity and thus not a direct effect of air temperature. Visual assessment of the map of predicted CO_2 concentrations revealed localized deviations from the expected pattern observed in adjacent sites. These discrepancies likely originated from errors in the catchment delineation, resulting in disparate predictions for seemingly similar stream sites. Employing a higher resolution DEM could have potentially mitigated these discrepancies and improved predictions in these specific areas. Within the stream network, summer CO_2 concentrations at lake outlets were substantially reduced, approaching the atmospheric saturation, due to higher residence time with air contact and primary production in lakes (Sand-Jensen, Riis & Kjær et al., 2022; Weyhenmeyer et al., 2012). The lake effect is pronounced downstream of the common eutrophic lakes in Denmark, where

Table 1
Summary Statistics of Estimated Stream CO_2 Fluxes (mmol m^{-2} d^{-1}) for Each Season

Season	$Q_{2.5\%}$	$Q_{50\%}$	$Q_{97.5\%}$	Mean
Autumn	118	237	593	268
Spring	89	191	492	218
Summer	115	225	529	253
Winter	118	241	651	276

Note. The fluxes were estimated from predicted CO_2 concentrations and empirical hydrological relationships.

21698961, 2024, 7, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024JG008031 by Royal Danish Library, Wiley Online Library on [07/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

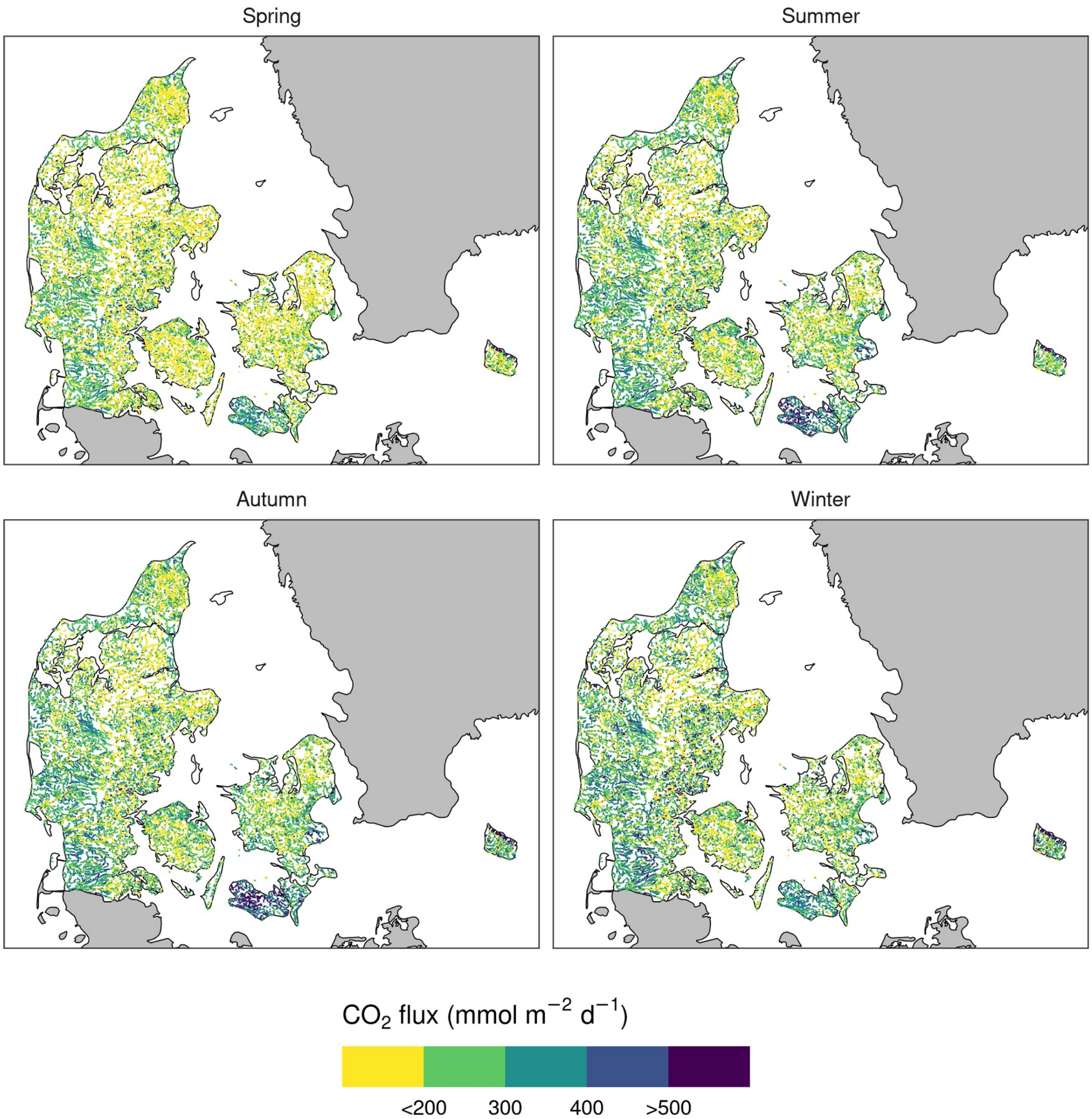

Figure 5. Estimated CO_2 flux for each season (a–d) covering Denmark. The fluxes are estimated from predicted CO_2 concentrations and empirical hydrological relationships. The stream network consists of sites included in the hydrological DK-model.

dissolved CO_2 essentially is consumed entirely, while the concentration increases with distance downstream from the lake outlet (Sand-Jensen, Riis & Kjær et al., 2022). Seasonal resolution of CO_2 concentrations is not only important for improving the accuracy of stream CO_2 fluxes but also for modeling performance and distribution of aquatic macrophytes whose rates of photosynthesis depend on CO_2 concentrations (Sand-Jensen, Riis & Martinsen et al., 2022; Demars & Trémolières, 2009).

Surprisingly, the hydrological variables were not among the most important predictors. We expected that DF would be important, as this flow component is expected to be very rich in dissolved CO_2 originating from soil respiration (Halbedel & Koschorreck, 2013; Marx et al., 2017; Sand-Jensen & Staehr, 2012). However, DF and OF components are more event-driven, in contrast to GWF which is more constant over time. The event-driven contributions may be neglected at the seasonal timescale applied in the analysis. Moreover, the flow components represent contributions resulting from processes simulated for a discrete location, which is very sensitive to the parameterization of the underlying hydrological model around that location, that is, conceptualization of hydrogeological layers or drain flow. Other predictors such as mean groundwater depth and HAND, which are

21698961, 2024, 7, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024JG008031 by Royal Danish Library, Wiley Online Library on [07/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

21698961, 2024, 7, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024JG008031 by Royal Danish Library, Wiley Online Library on [07/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

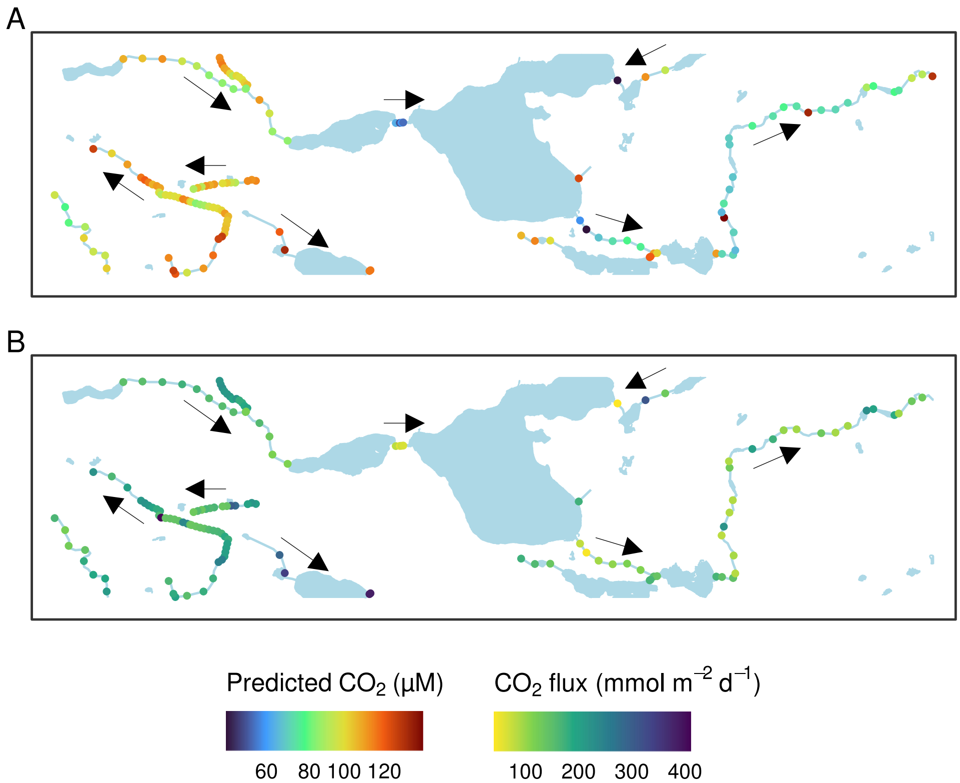

Figure 6. Predicted CO_2 concentration (a) and estimated fluxes (b) during summer around Lake Fure, north of Copenhagen. The geographical area is approximately 5 by 20 km with the stream network and lakes larger than 10^4 m^2 shown in blue. Three sites with high estimated CO_2 flux have been removed to improve visualization.

also expected to be correlated with water table depth (Koch et al., 2021; Rennó et al., 2008), are however important variables. Rocher-Ros et al. (2019) also found that a depth-to-water index improved predictions of stream CO_2 concentrations, highlighting the importance of soil-groundwater-stream interactions which can be highly discrete (Duvert et al., 2018; Lupon et al., 2019), calling for increased spatial sampling resolution. Similarly, catchment slope emerges as the most important predictor in agreement with several other studies (Lauerwald et al., 2015; Liu et al., 2022; Martinsen et al., 2020). Catchment slope may affect stream CO_2 concentrations in several ways but most importantly influences catchment carbon accumulation, that is, longer storage in low-slope landscapes, and higher CO_2 flux across the air-water interface (Smits et al., 2017; Wallin et al., 2011) indicating why slope repeatedly emerges as an essential proxy of stream CO_2 concentration.

4.2. Estimating and Measuring Stream CO_2 Fluxes

The relationship between estimated and in situ CO_2 fluxes is poor. While the spatial and temporal coverage of the in situ measurements is not high, they provide an initial evaluation of the estimated CO_2 fluxes which is often lacking in large-scale studies. While we did expect a better relationship, there is an apparent scale issue when comparing instantaneous and seasonal fluxes. The discrepancies appear to be larger during summer and similarly for test set predictions of CO_2 concentrations where the overestimated values are predominantly observations during spring. Spring and summer are seasons where ecosystem metabolism is highest (Kelly et al., 1983; Sand-Jensen & Frost-Christensen, 1998), and expectedly also the within-site variation. Consequently, diel variations in CO_2 concentrations can be pronounced (Sand-Jensen, Riis & Martinsen et al., 2022) complicating the comparison between instantaneous measurement and seasonally predicted CO_2 concentrations. Accounting for the type of primary producers, for example, benthic algae or submerged macrophytes, could improve future modeling efforts as reaches dominated by the latter appear to have prolonged periods of high production in contrast to that of benthic algae which are more temporally variable (Alnoee et al., 2016, 2020).

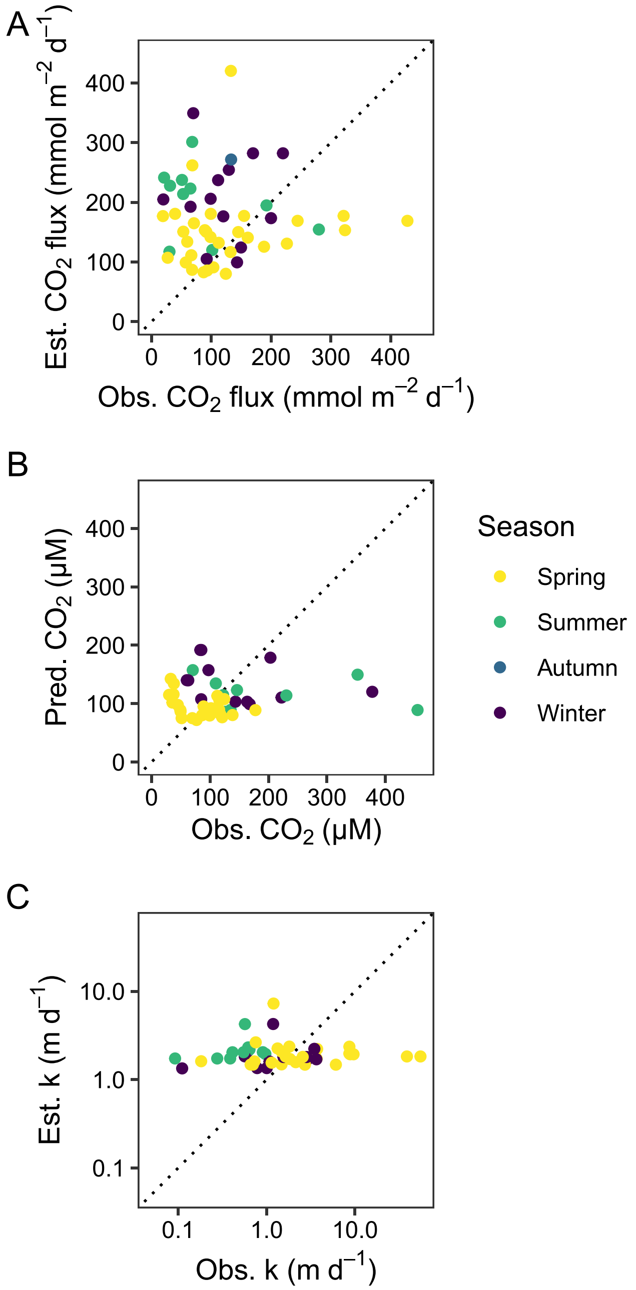

Figure 7. Estimated and observed stream CO_2 flux (a), CO_2 concentration (b), and k (c) colored by season and the 1:1 relationship (dotted line). Observed CO_2 concentrations were determined from pH and alkalinity, estimated k based on empirical relationships, and estimated CO_2 flux from estimated k and CO_2 concentration predicted by a Random Forest model.

In the comparison of estimated and in situ CO_2 fluxes, the overestimated fluxes appear to be a result of overestimated k values while CO_2 concentrations showed better agreement. Many of the overestimated fluxes were from the same sites, for example, six observations are from two sites close to each other in River Tryggevælde and three are from the same site in River Odense. This suggests that the estimated fluxes are not suitable for local (stream site) considerations but rather regional estimates of stream CO_2 fluxes. k is estimated in two steps using discharge and slope (Raymond et al., 2012) where the latter is more reliably determined from elevation models. However, the estimation of velocity solely based on discharge appears to introduce much higher uncertainty (Liu et al., 2022) which is also evident from the empirical relationship explaining a relatively low proportion of the variation ($R^2 = 0.49$). While aggregating multiple observations or empirical relationships may reduce the variability, there is a need for improved relationships to predict k in large-scale studies. Furthermore, the empirical relationship for k has a rather high intercept (2.02 m d^{-1}; Raymond et., 2,012). Lower k values have been reported in the literature (Sand-Jensen & Staehr, 2012; Wallin et al., 2011), suggesting a potential bias in upscaling, particularly in lowland regions like Denmark. Better knowledge of stream cross-sections or using hydrodynamic models that directly simulate velocity would likely improve estimates but might be difficult or computationally expensive to apply at scale. There has been recent progress in the development of empirical relationships, for example, Ulseth et al. (2019) identified different relationships for k in low- and high-energy streams but other variables might also cause differences in k or water velocity for otherwise similar streams including wind-exposure (Beaulieu et al., 2012), surface-films (Salter et al., 2011), and high submerged macrophyte biomass that markedly reduce mean velocity and introduce extensive small-scale variability of velocity within and outside macrophyte patches (Sand-Jensen, 2008; Sand-Jensen & Mebus, 1996). Developing relationships for k and differentiating between streams based on the most important drivers of k could further improve the models (Thyssen & Erlandsen, 1987; Wang et al., 2021), which is particularly important for small, lowland stream networks where CO_2 concentrations are high.

4.3. Large-Scale CO_2 Flux Estimates

The estimated average national stream CO_2 flux (3,675 g CO_2 m^{-2} stream y^{-1}) was of similar magnitude to those found in other studies when normalized by stream area. For Sweden, Humborg et al. (2010) estimated 6,785 g CO_2 m^{-2} y^{-1} and Wallin et al. (2018) estimated 14,104 g CO_2 m^{-2} y^{-1} only for low-order (1–4) streams and for the United States, Butman and Raymond (2011) estimated 8,690 g CO_2 m^{-2} y^{-1}.

The monitoring data used for analysis is generally collected during the daytime when CO_2 concentrations usually are lower due to photosynthesis compared to nighttime (Rocher-Ros et al., 2020; Sand-Jensen & Frost-Christensen, 1998) which could result in the underestimation of large-scale CO_2 fluxes. Multiple studies have found higher CO_2 fluxes in streams during nighttime than daytime (Attermeyer et al., 2021; Gómez-Gener et al., 2021). The degree of underestimation depends on the productivity of benthic primary producers and could thus be substantial in a Danish setting (Alnoee et al., 2020; Kelly et al., 1983; Sand-Jensen, Riis & Kjær et al., 2022).

Another essential parameter for upscaling fluxes is the stream area (Wallin et al., 2018). In our approach, we apply empirical relationships between upstream area and stream width, which does not consider seasonal variations. The smallest streams are not part of the stream network included in the DK-model. Our stream network correspond to approximately 29.5% of the length and 75% of the area compared to previously published data on stream

21698961, 2024, 7, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024JG008031 by Royal Danish Library, Wiley Online Library on [07/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

length and area for Denmark (Sand-Jensen et al., 2006). As smaller streams often have higher CO_2 concentrations and dominate the flux (Butman & Raymond, 2011; Wallin et al., 2018), the resulting large-scale estimates are underestimated. The increasing availability of high-resolution remote sensing imagery should enable improved stream area inventories in the future. While this can be challenging in certain habitats such as forests, better mapping is needed, particularly during the cold wet seasons when emissions are higher. Meanwhile, upscaling relies on empirical relationships as a function of the catchment area or discharge (Liu et al., 2022), which preferably should originate from the study region.

Driven by climate change, precipitation has increased (Pasten-Zapata et al., 2019) yielding higher groundwater levels in Denmark during the last decades bringing groundwater closer to the terrain and increasing the extent of flooding low-lying terrain (Seidenfaden et al., 2022). This development is expected to continue in the future, elevating the nationally averaged groundwater table by 0.12 m towards the end of the century (Seidenfaden et al., 2022). Individual climate models do however predict an average change of over 0.4 m with large spatial variations. This change is expected to increase the CO_2 flux from streams due to a greater influx of CO_2-rich groundwater and release from wider streams and periodically inundated terrestrial areas with easily degradable organic matter.

5. Conclusions

This study highlights the utility of machine learning models in predicting stream CO_2 concentrations and in turn estimating fluxes in Denmark. The final Random Forest model appeared to perform well, albeit with over-estimation at lower concentration levels but captured the expected seasonal patterns in the stream network. The most important predictors were catchment characteristics related to soil-groundwater-stream connectivity and seasonal variations. An apparent validation exercise using instantaneous in situ CO_2 flux measurements suggests future research efforts should aim at refining models at local scales and improve estimation of gas transfer velocity at larger scales. The study demonstrates the potential for predicting large-scale patterns in environmental variables such as CO_2 fluxes by integrating hydrological models and machine learning techniques. These findings have implications not only for ecosystem modeling but also for informing management strategies to adapt to stream carbon dynamics in a changing climate.

Data Availability Statement

Data used for the analysis are available from the sources cited in the main text. The DK-model data can be downloaded from https://hipdata.dk/. The flow components are available upon request, since they are not available via the data portal. The measurements of stream CO_2 fluxes, predicted CO_2 concentrations, estimated CO_2 fluxes, and scripts used for the analysis and figures are available from an online repository (https://doi.org/10.5281/zenodo.11072444; Martinsen et al., 2024).

Acknowledgments
This work was supported by a grant to Kaj Sand-Jensen from the Independent Research Fund Denmark (0217-00112B). We thank the two anonymous reviewers for their valuable comments, which improved the manuscript.

References

Abril, G., Bouillon, S., Darchambeau, F., Teodoru, C. R., Marwick, T. R., Tamooh, F., et al. (2015). Technical Note: Large overestimation of pCO_2 calculated from pH and alkalinity in acidic, organic-rich freshwaters. *Biogeosciences*, *12*(1), 67–78. https://doi.org/10.5194/bg-12-67-2015

Adhikari, K., Kheir, R. B., Greve, M. B., Bøcher, P. K., Malone, B. P., Minasny, B., et al. (2013). High-resolution 3-D mapping of soil texture in Denmark. *Soil Science Society of America Journal*, *77*(3), 860–876. https://doi.org/10.2136/sssaj2012.0275

Alnoee, A. B., Levi, P. S., Baattrup-Pedersen, A., & Riis, T. (2020). Macrophytes enhance reach-scale metabolism on a daily, seasonal and annual basis in agricultural lowland streams. *Aquatic Sciences*, *83*(1), 11. https://doi.org/10.1007/s00027-020-00766-4

Alnoee, A. B., Riis, T., & Baattrup-Pedersen, A. (2016). Comparison of metabolic rates among macrophyte and nonmacrophyte habitats in streams. *Freshwater Science*, *35*(3), 834–844. https://doi.org/10.1086/687842

Attermeyer, K., Casas-Ruiz, J. P., Fuss, T., Pastor, A., Cauvy-Fraunié, S., Sheath, D., et al. (2021). Carbon dioxide fluxes increase from day to night across European streams. *Communications Earth & Environment*, *2*(1), 1–8. https://doi.org/10.1038/s43247-021-00192-w

Bastviken, D., Sundgren, I., Natchimuthu, S., Reyier, H., & Gålfalk, M. (2015). Technical Note: Cost-efficient approaches to measure carbon dioxide (CO_2) fluxes and concentrations in terrestrial and aquatic environments using mini loggers. *Biogeosciences*, *12*(12), 3849–3859. https://doi.org/10.5194/bg-12-3849-2015

Beaulieu, J. J., Shuster, W. D., & Rebholz, J. A. (2012). Controls on gas transfer velocities in a large river. *Journal of Geophysical Research*, *117*(G2). https://doi.org/10.1029/2011JG001794

Breiman, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science*, *16*(3), 199–231. https://doi.org/10.1214/ss/1009213726

Butman, D., & Raymond, P. A. (2011). Significant efflux of carbon dioxide from streams and rivers in the United States. *Nature Geoscience*, *4*(12), 839–842. https://doi.org/10.1038/ngeo1294

21698961, 2024, 7, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024JG008031 by Royal Danish Library, Wiley Online Library on [07/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Butman, D., Stackpoole, S., Stets, E., McDonald, C. P., Clow, D. W., & Striegl, R. G. (2016). Aquatic carbon cycling in the conterminous United States and implications for terrestrial carbon accounting. *Proceedings of the National Academy of Sciences*, *113*(1), 58–63. https://doi.org/10.1073/pnas.1512651112

Cole, J. J., Prairie, Y. T., Caraco, N. F., McDowell, W. H., Tranvik, L. J., Striegl, R. G., et al. (2007). Plumbing the global carbon cycle: Integrating inland waters into the terrestrial carbon budget. *Ecosystems*, *10*(1), 172–185. https://doi.org/10.1007/s10021-006-9013-8

Crawford, J. T., Lottig, N. R., Stanley, E. H., Walker, J. F., Hanson, P. C., Finlay, J. C., & Striegl, R. G. (2014). CO_2 and CH_4 emissions from streams in a lake-rich landscape: Patterns, controls, and regional significance. *Global Biogeochemical Cycles*, *28*(3), 197–210. https://doi.org/10.1002/2013GB004661

Demars, B. O., & Trémolières, M. (2009). Aquatic macrophytes as bioindicators of carbon dioxide in groundwater fed rivers. *Science of the Total Environment*, *407*(16), 4752–4763. https://doi.org/10.1016/j.scitotenv.2009.04.017

Duvert, C., Butman, D. E., Marx, A., Ribolzi, O., & Hutley, L. B. (2018). CO_2 evasion along streams driven by groundwater inputs and geomorphic controls. *Nature Geoscience*, *11*(11), 813–818. https://doi.org/10.1038/s41561-018-0245-y

Gómez-Gener, L., Rocher-Ros, G., Battin, T., Cohen, M. J., Dalmagro, H. J., Dinsmore, K. J., et al. (2021). Global carbon dioxide efflux from rivers enhanced by high nocturnal emissions. *Nature Geoscience*, *14*(5), 289–294. https://doi.org/10.1038/s41561-021-00722-3

Gran, G. (1952). Determination of the equivalence point in potentiometric titrations. Part II. *Analyst*, *77*(920), 661–671. https://doi.org/10.1039/an9527700661

Halbedel, S., & Koschorreck, M. (2013). Regulation of CO_2 emissions from temperate streams and reservoirs. *Biogeosciences*, *10*(11), 7539–7551. https://doi.org/10.5194/bg-10-7539-2013

Hastie, A., Lauerwald, R., Weyhenmeyer, G., Sobek, S., Verpoorter, C., & Regnier, P. (2018). CO_2 evasion from boreal lakes: Revised estimate, drivers of spatial variability, and future projections. *Global Change Biology*, *24*(2), 711–728. https://doi.org/10.1111/gcb.13902

Henriksen, H. J., Schneider, R., Koch, J., Ondracek, M., Troldborg, L., Seidenfaden, I. K., et al. (2023). A new digital twin for climate change adaptation, water management, and disaster risk reduction (HIP digital twin). *Water*, *15*(1), 25. https://doi.org/10.3390/w15010025

Hofmann, A. F., Soetaert, K., Middelburg, J. J., & Meysman, F. J. R. (2010). AquaEnv - an aquatic acid-base modelling environment in R. *Aquatic Geochemistry*, *16*(4), 507–546. https://doi.org/10.1007/s10498-009-9084-1

Højberg, A. L., Troldborg, L., Stisen, S., Christensen, B. B. S., & Henriksen, H. J. (2013). Stakeholder driven update and improvement of a national water resources model. *Environmental Modelling and Software*, *40*, 202–213. https://doi.org/10.1016/j.envsoft.2012.09.010

Hotchkiss, E. R., Hall Jr., R. O., Sponseller, R. A., Butman, D., Klaminder, J., Laudon, H., et al. (2015). Sources of and processes controlling CO_2 emissions change with the size of streams and rivers. *Nature Geoscience*, *8*(9), 696–699. https://doi.org/10.1038/ngeo2507

Humborg, C., Mörth, C.-M., Sundbom, M., Borg, H., Blenckner, T., Giesler, R., & Ittekkot, V. (2010). CO_2 supersaturation along the aquatic conduit in Swedish watersheds as constrained by terrestrial respiration, aquatic respiration and weathering. *Global Change Biology*, *16*(7), 1966–1978. https://doi.org/10.1111/j.1365-2486.2009.02092.x

Hutchins, R. H. S., Prairie, Y. T., & Giorgio, P. A. D. (2019). Large-scale landscape drivers of CO_2, CH_4, DOC, and DIC in boreal river networks. *Global Biogeochemical Cycles*, *33*(2), 125–142. https://doi.org/10.1029/2018GB006106

Jähne, B., Heinz, G., & Dietrich, W. (1987). Measurement of the diffusion coefficients of sparingly soluble gases in water. *Journal of Geophysical Research*, *92*(C10), 10767–10776. https://doi.org/10.1029/jc092ic10p10767

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 103). Springer. https://doi.org/10.1007/978-1-4614-7138-7

Jones, J. B., & Mulholland, P. J. (1998). Influence of drainage basin topography and elevation on carbon dioxide and methane supersaturation of stream water. *Biogeochemistry*, *40*(1), 57–72. https://doi.org/10.1023/A:1005914121280

Kelly, M. G., Thyssen, N., & Moeslund, B. (1983). Light and the annual variation of oxygen- and carbon-based measurements of productivity in a macrophyte-dominated river1. *Limnology & Oceanography*, *28*(3), 503–515. https://doi.org/10.4319/lo.1983.28.3.0503

Koch, J., Gotfredsen, J., Schneider, R., Troldborg, L., Stisen, S., & Henriksen, H. J. (2021). High resolution water table modeling of the shallow groundwater using a knowledge-guided gradient boosting decision tree model. *Frontiers in Water*, *3*, 701726. https://doi.org/10.3389/frwa.2021.701726

Lauerwald, R., Laruelle, G. G., Hartmann, J., Ciais, P., & Regnier, P. A. G. (2015). Spatial patterns in CO_2 evasion from the global river network. *Global Biogeochemical Cycles*, *29*(5), 534–554. https://doi.org/10.1002/2014gb004941

Lauridsen, T. L., Søndergaard, M., Jensen, J. P., & Jeppesen, E. (2005). *Investigations in lakes (in Danish)* (p. 234). Danish Center for Environment. Retrieved from https://odaforalle.au.dk/

Levin, G. (2022). *Basemap04. Documentation of the data and method for elaboration of a land use and land cover map for Denmark*. DCE – Danish Centre for Environment and Energy: Aarhus University. Retrieved from http://dce2.au.dk/pub/TR252.pdf

Lindsay, J. B. (2016a). Whitebox GAT: A case study in geomorphometric analysis. *Computers and Geosciences*, *95*, 75–84. https://doi.org/10.1016/j.cageo.2016.07.003

Lindsay, J. B. (2016b). Efficient hybrid breaching-filling sink removal methods for flow path enforcement in digital elevation models. *Hydrological Processes*, *30*(6), 846–857. https://doi.org/10.1002/hyp.10648

Liu, S., Kuhn, C., Amatulli, G., Aho, K., Butman, D. E., Allen, G. H., et al. (2022). The importance of hydrology in routing terrestrial carbon to the atmosphere via global streams and rivers. *Proceedings of the National Academy of Sciences*, *119*(11), e2106322119. https://doi.org/10.1073/pnas.2106322119

Liu, S., & Raymond, P. A. (2018). Hydrologic controls on pCO_2 and CO_2 efflux in US streams and rivers. *Limnology and Oceanography Letters*, *3*(6), 428–435. https://doi.org/10.1002/lol2.10095

Long, H., Vihermaa, L., Waldron, S., Hoey, T., Quemin, S., & Newton, J. (2015). Hydraulics are a first-order control on CO2 efflux from fluvial systems. *Journal of Geophysical Research: Biogeosciences*, *120*(10), 1912–1922. https://doi.org/10.1002/2015JG002955

Lupon, A., Denfeld, B. A., Laudon, H., Leach, J., Karlsson, J., & Sponseller, R. A. (2019). Groundwater inflows control patterns and sources of greenhouse gas emissions from streams. *Limnology & Oceanography*, *64*(4), 1545–1557. https://doi.org/10.1002/lno.11134

Martinsen, K. T., Kragh, T., & Sand-Jensen, K. (2020). Carbon dioxide partial pressure and emission throughout the Scandinavian stream network. *Global Biogeochemical Cycles*, *34*(12), e2020GB006703. https://doi.org/10.1029/2020GB006703

Martinsen, K. T., Sand-Jensen, K., Bergmann, V., Skjærlund, T., Kjær, J. E., & Koch, J. (2024). Seasonal carbon dioxide concentrations and fluxes throughout Denmark's stream network [Dataset]. *Zenodo*. https://doi.org/10.5281/zenodo.11072444

Marx, A., Dusek, J., Jankovec, J., Sanda, M., Vogel, T., van Geldern, R., et al. (2017). A review of CO_2 and associated carbon dynamics in headwater streams: A global perspective. *Reviews of Geophysics*, *55*(2), 560–585. https://doi.org/10.1002/2016rg000547

O'Callaghan, J. F., & Mark, D. M. (1984). The extraction of drainage networks from digital elevation data. *Computer Vision, Graphics, and Image Processing*, *28*(3), 323–344. https://doi.org/10.1016/S0734-189X(84)80011-0

21698961, 2024, 7, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024JG008031 by Royal Danish Library, Wiley Online Library on [07/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Olden, J. D., Lawler, J. J., & Poff, N. L. (2008). Machine learning methods without tears: A primer for ecologists. *The Quarterly Review of Biology*, *83*(2), 171–193. https://doi.org/10.1086/587826

Olsen, M., & Højberg, A. L. (2011). *Expansion of stream network (in Danish)*. Geological Survey of Denmark and Greenland (GEUS).

Pasten-Zapata, E., Sonnenborg, T. O., & Refsgaard, J. C. (2019). Climate change: Sources of uncertainty in precipitation and temperature projections for Denmark. *GEUS Bulletin*, *43*. https://doi.org/10.34194/GEUSB-201943-01-02

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Raymond, P. A., Hartmann, J., Lauerwald, R., Sobek, S., McDonald, C., Hoover, M., et al. (2013). Global carbon dioxide emissions from inland waters. *Nature*, *503*(7476), 355–359. https://doi.org/10.1038/nature12760

Raymond, P. A., Zappa, C. J., Butman, D., Bott, T. L., Potter, J., Mulholland, P., et al. (2012). Scaling the gas transfer velocity and hydraulic geometry in streams and small rivers. *Limnology and Oceanography: Fluids and Environments*, *2*(1), 41–53. https://doi.org/10.1215/21573689-1597669

Rebsdorf, A., Thyssen, N., & Erlandsen, M. (1991). Regional and temporal variation in pH, alkalinity and carbon dioxide in Danish streams, related to soil type and land use. *Freshwater Biology*, *25*(3), 419–435. https://doi.org/10.1111/j.1365-2427.1991.tb01386.x

Rennó, C. D., Nobre, A. D., Cuartas, L. A., Soares, J. V., Hodnett, M. G., Tomasella, J., & Waterloo, M. J. (2008). HAND, a new terrain descriptor using SRTM-DEM: Mapping terra-firme rainforest environments in Amazonia. *Remote Sensing of Environment*, *112*(9), 3469–3481. https://doi.org/10.1016/j.rse.2008.03.018

Rocher-Ros, G., Sponseller, R. A., Bergström, A.-K., Myrstener, M., & Giesler, R. (2020). Stream metabolism controls diel patterns and evasion of CO_2 in Arctic streams. *Global Change Biology*, *26*(3), 1400–1413. https://doi.org/10.1111/gcb.14895

Rocher-Ros, G., Sponseller, R. A., Lidberg, W., Mörth, C.-M., & Giesler, R. (2019). Landscape process domains drive patterns of CO_2 evasion from river networks. *Limnology and Oceanography Letters*, *4*(4), 87–95. https://doi.org/10.1002/lol2.10108

Salter, M. E., Upstill-Goddard, R. C., Nightingale, P. D., Archer, S. D., Blomquist, B., Ho, D. T., et al. (2011). Impact of an artificial surfactant release on air-sea gas fluxes during deep ocean gas exchange experiment II. *Journal of Geophysical Research*, *116*(C11). https://doi.org/10.1029/2011JC007023

Sand-Jensen, K. (2008). Drag forces on common plant species in temperate streams: Consequences of morphology, velocity and biomass. *Hydrobiologia*, *610*(1), 307–319. https://doi.org/10.1007/s10750-008-9446-5

Sand-Jensen, K., Friberg, N., & Murphy, J. (Eds.) (2006)., *Running Waters: Historical development and restoration of lowland Danish streams*. Aarhus Universitetsforlag.

Sand-Jensen, K., & Frost-Christensen, H. (1998). Photosynthesis of amphibious and obligately submerged plants in CO_2-rich lowland streams. *Oecologia*, *117*(1–2), 31–39. https://doi.org/10.1007/s004420050628

Sand-Jensen, K., & Mebus, J. R. (1996). Fine-scale patterns of water velocity within macrophyte patches in streams. *Oikos*, *76*(1), 169–180. https://doi.org/10.2307/3545759

Sand-Jensen, K., Riis, T., Kjær, J. E., & Martinsen, K. T. (2022a). Stream-Lake connectivity is an important control of fluvial CO_2 concentrations and emissions in catchments. *Earth and Space Science*, *9*(12), e2022EA002664. https://doi.org/10.1029/2022EA002664

Sand-Jensen, K., Riis, T., & Martinsen, K. T. (2022b). Photosynthesis, growth, and distribution of plants in lowland streams—A synthesis and new data analyses of 40 years research. *Freshwater Biology*, *67*(7), 1255–1271. https://doi.org/10.1111/fwb.13915

Sand-Jensen, K., & Staehr, P. A. (2012). CO_2 dynamics along Danish lowland streams: Water-air gradients, piston velocities and evasion rates. *Biogeochemistry*, *111*(1–3), 615–628. https://doi.org/10.1007/s10533-011-9696-6

Scharling, M. (1999a). *Climate grid Denmark: Precipitation 10*10 km (ver. 2) (in Danish)* (Vol. 99). Danish Meteorological Institute.

Scharling, M. (1999b). *Climate grid Denmark: Precipitation, air temperature, and potential evapotranspiration 20x20 and 40x40 km: Method description (in Danish)*. Danish Meteorological Institute.

SDFI. (2021). Agency for datasupply and infrastructure, SDFI. Retrieved from https://dataforsyningen.dk/data

Seidenfaden, I. K., Sonnenborg, T. O., Stisen, S., & Kidmose, J. (2022). Quantification of climate change sensitivity of shallow and deep groundwater in Denmark. *Journal of Hydrology: Regional Studies*, *41*, 101100. https://doi.org/10.1016/j.ejrh.2022.101100

Smits, A. P., Schindler, D. E., Holtgrieve, G. W., Jankowski, K. J., & French, D. W. (2017). Watershed geomorphology interacts with precipitation to influence the magnitude and source of CO_2 emissions from Alaskan streams. *Journal of Geophysical Research: Biogeosciences*, *122*(8), 1903–1921. https://doi.org/10.1002/2017jg003792

Soltani, M., Bjerre, E., Koch, J., & Stisen, S. (2021). Integrating remote sensing data in optimization of a national water resources model to improve the spatial pattern performance of evapotranspiration. *Journal of Hydrology*, *603*, 127026. https://doi.org/10.1016/j.jhydrol.2021.127026

Tarboton, D. G. (2017). *Terrain analysis using digital elevation models (TauDEM) Utah Water Research Laboratory*. Utah State University. Retrieved from http://hydrology.usu.edu/taudem/taudem5/index.html

Thyssen, N., & Erlandsen, M. (1987). Reaeration of oxygen in shallow, macrophyte rich streams: II. Relationship between the reaeration rate coefficient and hydraulic properties. *Internationale Revue Der Gesamten Hydrobiologie Und Hydrographie*, *72*(5), 575–597. https://doi.org/10.1002/iroh.19870720505

Ulseth, A. J., Hall, R. O., Boix Canadell, M., Madinger, H. L., Niayifar, A., & Battin, T. J. (2019). Distinct air–water gas exchange regimes in low- and high-energy streams. *Nature Geoscience*, *12*(4), 259–263. https://doi.org/10.1038/s41561-019-0324-8

Wallin, M. B., Campeau, A., Audet, J., Bastviken, D., Bishop, K., Kokic, J., et al. (2018). Carbon dioxide and methane emissions of Swedish low-order streams - a national estimate and lessons learnt from more than a decade of observations. *Limnology and Oceanography Letters*, *3*(3), 156–167. https://doi.org/10.1002/lol2.10061

Wallin, M. B., Grabs, T., Buffam, I., Laudon, H., Ågren, A., Öquist, M. G., & Bishop, K. (2013). Evasion of CO_2 from streams–The dominant component of the carbon export through the aquatic conduit in a boreal landscape. *Global Change Biology*, *19*(3), 785–797. https://doi.org/10.1111/gcb.12083

Wallin, M. B., Öquist, M. G., Buffam, I., Billett, M. F., Nisell, J., & Bishop, K. H. (2011). Spatiotemporal variability of the gas transfer coefficient (KCO2) in boreal streams: Implications for large scale estimates of CO2 evasion. *Global Biogeochemical Cycles*, *25*(3). https://doi.org/10.1029/2010gb003975

Wang, J., Bombardelli, F. A., & Dong, X. (2021). Physically based scaling models to predict gas transfer velocity in streams and rivers. *Water Resources Research*, *57*(3), e2020WR028757. https://doi.org/10.1029/2020WR028757

Wanninkhof, R. (1992). Relationship between wind speed and gas exchange over the ocean. *Journal of Geophysical Research*, *97*(C5), 7373–7382. https://doi.org/10.1029/92jc00188

Weyhenmeyer, G. A., Kortelainen, P., Sobek, S., Müller, R., & Rantakari, M. (2012). Carbon dioxide in boreal surface waters: A comparison of lakes and streams. *Ecosystems*, *15*(8), 1295–1307. https://doi.org/10.1007/s10021-012-9585-4

21698961, 2024, 7, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024JG008031 by Royal Danish Library, Wiley Online Library on [07/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License