Check for updates

LIMNOLOGY and OCEANOGRAPHY

ASLO

Limnol. Oceanogr. 70, 2025, 1152–1166
© 2025 The Author(s). *Limnology and Oceanography* published by Wiley Periodicals LLC on behalf of Association for the Sciences of Limnology and Oceanography.
doi: 10.1002/lno.70019

RESEARCH ARTICLE

Predicting river ecosystem metabolism across large environmental gradients: Drivers and temporal dependencies in the Iberian Peninsula

Amaia A. Rodeles [1,2*] **Francisco J. Peñas** [1] **Maite Arroita** [3] **José Barquín** [1]

[1]IHCantabria—Instituto de Hidráulica Ambiental de la Universidad de Cantabria, Santander, Spain; [2]Centro de Investigación Mariña (CIM), Universidade de Vigo, Vigo, 36310, Spain; [3]Department of Plant Biology and Ecology, University of the Basque Country, Bilbao, Spain

Abstract

River ecosystem metabolism plays a significant role in the global carbon cycle. However, the limited spatial or temporal scale of most river metabolism studies hinders our ability to draw general patterns, identify common drivers, and make reliable global predictions. We developed Random Forest models for predicting daily metabolism rates using a large database of more than 100 river reaches across the Iberian Peninsula covering a large environmental gradient. As potential drivers, we included static variables (e.g., catchment area, distance to the sea), anthropogenic factors (e.g., land uses), and short-term dynamic variables (e.g., light, water temperature, discharge) averaged over different periods (from 0 to 40 d) to explore the role of shorter vs. longer-term environmental control on daily river metabolism rates. Both daily gross primary production and ecosystem respiration rates responded more strongly to average environmental conditions over the previous 40 d than to daily values. The 40-d average random forest models explained up to 77% of gross primary production and 82% of ecosystem respiration variance. The most important drivers of GPP were stage (depth), distance to the sea, and light, while the main predictors of ER were stage and GPP. Dynamic variables were generally the most important drivers of daily metabolic rates, although static ones such as distance to the sea also played a role. Our results indicate that temporal patterns in river metabolism are influenced by a combination of environmental conditions integrated over several weeks, seasonal timing, and to a lesser extent, topology.

Ecosystem metabolism integrates the flow of matter and energy across ecosystem components via the assimilation of inorganic carbon through photosynthesis (gross primary production [GPP]) and the mineralization of organic carbon through respiration (ecosystem respiration [ER]). Metabolism rates in terrestrial and oceanic ecosystems have been comprehensively studied, as these ecosystems are important carbon sinks that could mitigate climate change (Field et al. 1998; Allen et al. 2005; Boscolo-Galazzo et al. 2018). In contrast, the contribution of river ecosystem metabolism to the global carbon cycle was considered negligible and was thus overlooked in the past, although more recent research has challenged this view (Cole et al. 2007; Raymond et al. 2013; Battin et al. 2023). Nevertheless, studies integrating river metabolism with its terrestrial and oceanic counterparts are still scarce due to the lack of accurate global river metabolism estimations. The difficulty of resolving regional and global patterns and drivers of metabolism across biomes may be rooted in two main causes: (1) the reliance of river metabolism estimation upon local field measurements and (2) the complexity of potentially interacting factors that may influence river metabolism at broad scales (Bernhardt et al. 2018).

While remote sensing data have been applied to quantify global terrestrial and oceanic productivity (Potter et al. 1993; Buitenhuis et al. 2013; Bai et al. 2020), river metabolism estimates are reliant on in situ continuous monitoring (e.g., dissolved oxygen concentrations). Moreover, in the past, the high cost of oxygen sensors and their maintenance constrained the spatial and temporal scale of river ecosystem metabolism studies from a few days to a few weeks under good

*Correspondence: amaia.angulo@uvigo.es

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

Associate editor: Ryan Sponseller

Data Availability Statement: All study site information, metabolism model inputs, and model code are available at the following Figshare data repository: https://doi.org/10.6084/m9.figshare.c.7581881

environmental conditions (Finlay 2011; Alnoee et al. 2015; Bernhardt et al. 2018). Currently, the commercialization of cheaper and more durable sensors, the installation of continuous monitoring stations, and the development of new analytical tools, together with an increase in computing power, are opening the door to long-term and large spatial-scale studies (Appling, Read, et al. 2018; Savoy et al. 2019; Bernhardt et al. 2022). However, the accurate extrapolation of these results to other rivers in which field measurements are not available remains difficult due to the complex spatiotemporal interactions between river metabolism drivers.

Despite the technical limitations of river metabolism modeling, significant advances have been made in recent decades in the characterization of river metabolic patterns and in the description of the key drivers using local studies. These achievements have allowed for the development of simple conceptual frameworks that define the most important direct and indirect factors that shape metabolic rates, such as incident light, hydrology, and water chemistry, but also a range of land-use activities (Bernot et al. 2010; Bernhardt et al. 2022). Accordingly, light and flow are thought to be the most important factors driving the spatiotemporal variability of GPP (Uehlinger et al. 2003; Roberts et al. 2007; Izagirre et al. 2008), whereas ER seems mainly related to water temperature, nutrient concentrations, and organic matter quantity and quality (Young and Huryn 1999; Izagirre et al. 2008; Segatto et al. 2021). Regional drivers such as topography and catchment area may also affect river ecosystem metabolism through their effect on local variables such as discharge and temperature (Finlay 2011; Savoy et al. 2019). Other local variables, such as canopy cover, may also regulate river metabolic rates in small streams by limiting light reaching the water surface and thereby reducing GPP rates (Bernhardt et al. 2018; Koenig et al. 2019), and by influencing ER via allochthonous organic matter inputs (Roberts et al. 2007; Griffiths et al. 2013; Bernhardt et al. 2018). Moreover, river metabolism also integrates anthropogenic impacts within the broader catchment, which may interact in non-linear ways, including hydrological alterations (Genzoli and Hall 2016; Arroita et al. 2017; Chowanski et al. 2020), land-use changes (Young and Huryn 1999; Bernot et al. 2010; Fuß et al. 2017), pollution (Kosinski 1984; Aristi et al. 2016; Arroita et al. 2019), and climate change (Boix Canadell et al. 2021). Thus, unraveling the interplay among possible factors becomes a rather complicated endeavor.

Despite this progress, river metabolism studies, particularly those capturing longer time series, have mostly focused on temperate streams and rivers in the Northern Hemisphere and thus are unlikely to include the diversity of metabolic regimes and environmental conditions that occur worldwide. Therefore, using long time-series of data (at least a year of daily metabolic rates) covering a wide spectrum of environmental conditions, including rivers of varied sizes and different biomes, represents an important next step in advancing our understanding of the processes driving river ecosystems from local to regional and larger scales (Rodríguez-Castillo et al. 2019; Savoy et al. 2019; Bernhardt et al. 2022). Toward this goal, we present and analyze long-term time series of metabolism data from rivers that encompass a diverse range of underexplored topographical and climatic conditions across the Iberian Peninsula in Southern Europe.

Our assessment of daily metabolic rates follows a general conceptual framework for river ecosystem metabolism prediction where static, regional characteristics (catchment area, topography, etc.), climate, and anthropogenic impacts (land use, pollution, hydrological alteration, etc.) shape metabolic rates by regulating direct, dynamic factors (discharge, light, water temperature, etc.). These dynamic variables, in turn, control daily river GPP and ER, with primary production also regulating respiration (Fig. 1). Additionally, our framework also includes the temporal scales at which these dynamic variables influence daily metabolic rates. With some exceptions (Rodríguez-Castillo 2017), studies of daily river metabolism use same-day dynamic variables for driver analysis (light, discharge, temperature, etc.). Therefore, in those analyses, metabolic rates are only related to immediate environmental changes, which may reflect instantaneous biological responses. Yet, river metabolism integrates the biological activity of all organisms (including multicellular organisms), which may also be regulated by processes of growth and senescence affected by longer-term environmental conditions and density-dependent and stochastic factors (Munn et al. 2020; Blaszczak et al. 2023). Studying the effect of longer-term (days to weeks) environmental dynamics on daily metabolic rates may improve predictions and expose the importance of the temporal dimension of river metabolism.

The present study aims at (1) generating a model to accurately predict daily GPP and ER rates using dynamic, regional, and anthropogenic drivers across large environmental gradients to broaden our understanding of river metabolic regimes and their controls; (2) identifying the main environmental drivers of daily river ecosystem metabolism rates; (3) analyzing the effect of the temporal inertia of dynamic environmental variables (discharge, stage, light, water temperature, and water chemistry) in daily river metabolic rates by integrating their effect over different time periods. To achieve this, we used an extensive monitoring network in Spain, which incorporates more than 100 stations located in almost every large catchment of the country. The Iberian Peninsula is a perfect setting to analyze river ecosystem functioning patterns and their main drivers, as well as to train statistical models, as it encompasses a large gradient of physical, chemical, and hydrological conditions reflective of major transitions across temperate, Mediterranean, and semi-arid climates (Peñas and Barquín 2019). We hypothesize that the main factors controlling spatiotemporal patterns of daily river ecosystem metabolism would be the dynamic variables, while static regional and anthropogenic factors would show weaker effects due to their indirect control over metabolism (Fig. 1). Moreover, we expect that GPP would be associated with shorter time lapses of the

19395590, 2025, 5, Downloaded from https://aslopubs.onlinelibrary.wiley.com/doi/10.1002/lno.70019 by Universidad De Cantabria, Wiley Online Library on [04/07/2025]. 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

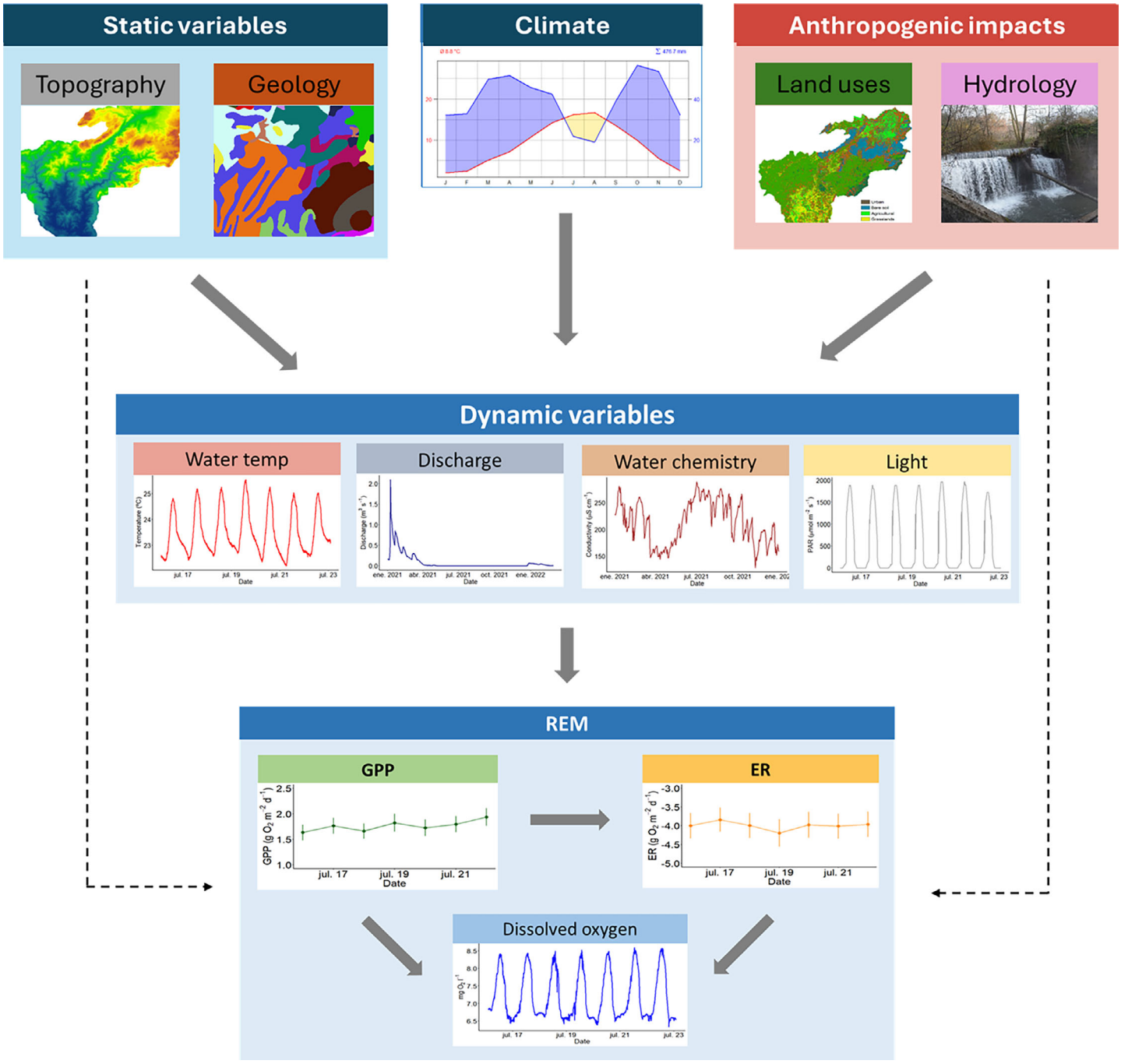

Fig. 1. Conceptual framework of the main drivers of river ecosystem metabolism based on current knowledge. Static regional variables (land topography, river topology, geology, etc.), climate and anthropogenic impacts (land uses, hydrological alteration, pollution, etc.) directly regulate short-term dynamic variables (water temperature, discharge, light availability, and water chemistry). Then, these dynamic variables control daily rates by regulating biological activity at different temporal scales. Moreover, regional static and anthropogenic impacts may also weakly affect river metabolism directly.

dynamic variables, as it depends on instantaneous light availability and photosynthetic organisms, many of them characterized by rapid growth rates. In contrast, ER is integrated by community composition and the relative abundance of groups of organisms with highly variable response times to changing environmental conditions, from microbes (quick responses) to fishes (slow responses). Moreover, ER is also governed by the decomposition rates of allochthonous organic matter, which

19395590, 2025, 5, Downloaded from https://aslopubs.onlinelibrary.wiley.com/doi/10.1002/lno.70019 by Universidad De Cantabria, Wiley Online Library on [04/07/2025]. 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

presents highly variable biological reactivity depending on its origin (Bertuzzo et al. 2022). Thus, we expect ER to be controlled by dynamic variables integrated over longer periods.

Methods

Study area

The study area covers continental Spain, an area of more than 0.5 million km^2. Spain is located in southwest Europe, surrounded by the Cantabrian Sea to the North, the Atlantic Ocean to the West, and the Mediterranean Sea to the east. The dominant climate is Mediterranean, with dry and hot summers and mild and humid winters; although a small northern area bordering the Cantabrian Sea and Atlantic Ocean presents a temperate oceanic climate, with mild temperatures and high precipitation all year long. The complex topography of Spain, with a high plateau in the center surrounded and crossed by various mountain ranges, as well as the clash of climates, results in a wide variety of bioclimatic conditions and ecosystems, from broadleaf forests in the north to landscapes dominated by shrubs and deserts in the southeast (Rivas-Martínez 1987). Consequently, rivers present high flow variability due to these precipitation and temperature regimes, with some rivers having a marked seasonality (Peñas and Barquín 2019). The main river basins are, north to south: Miño, Duero, Tajo, Guadiana, and Guadalquivir draining into the Atlantic Ocean and Ebro and Júcar draining into the Mediterranean Sea. In addition, there is an important group of smaller catchments draining the Cantabrian cordillera to the north (Fig. 2).

Time series acquisition and data preparation

The Spanish Automatic System of Water Quality Information (SAICA for its Spanish abbreviation) is a network of 175 continuous monitoring stations (Fig. 2) managed by the Government of Spain through the river basin authorities established in the 2000s with the objective of monitoring water quality characteristics in critical river reaches for water management purposes (Ministry for the Ecological Transition and the Demographic Challenge 2020). Thus, the location of SAICA stations was based on the level of anthropogenic pressure in the river reach (mainly water extraction and pollution). SAICA stations (from now on "sites") provide continuous data for some of the main variables needed for river ecosystem metabolism estimation (dissolved oxygen, water temperature, and discharge) at 15-min intervals. We discarded sites located too close to upstream large dams (< 5 km) as reservoirs may affect dissolved oxygen concentrations, violating modeling assumptions. When discharge data were absent, we used discharge data from the nearest gauge station (Center for Study and Experimentation of Construction Works 2020) corrected by catchment area (Supporting Information Fig. S2A). After dissolved oxygen modeling tests, some results showed unrealistic metabolic estimates (negative GPP or positive ER) due to high discharge; so, to reduce noise, we removed days with discharge higher than three times the median of the site for the study period (Supporting Information Fig. S2D). After this selection process, we kept 106 sites out of the original 175 (Fig. 2). The selected sites were located in rivers of stream order equal to or greater than 5.

Accurate hydrological depth estimations were not available, and due to a lack of data and high hydromorphological variability both between sites and within each site during the study period, it could not be accurately assessed. Therefore, we used stage as a proxy for water depth, as its relationship with discharge was site-specific and the error at each site was generally constant, thus resulting in a more confident estimation of intra-site river metabolism variability (for a detailed explanation of our reasoning, see Supporting Information Heading One). To fill in the gaps where stage or discharge were missing, we fitted site-specific power equations (Leopold and Maddock 1953) using the available stage and discharge data at each site and the R package *aomisc* (Supporting Information Fig. S2B). Gaps up to 3 h in dissolved oxygen and water temperature were linearly interpolated to 15-min intervals as in Appling, Read, et al. (2018) using the *na.approx* function of the *zoo* package (Zeileis et al. 2021) in R 4.0.2 (Supporting Information Fig. S2C).

Dissolved oxygen models to estimate river ecosystem metabolism also required two other variables not provided by SAICA: atmospheric pressure and surface solar radiation downward, which were needed to calculate the theoretical saturation concentration of oxygen and photosynthetic photon flux density (PPFD). We obtained hourly atmospheric pressure (Pa) and surface solar radiation downward (J m^{-2}) data series from the Copernicus Climate Change Service (Muñoz Sabater 2019), which provides reanalysis of remote sensing data at a raster cell resolution of 9 km (Supporting Information Fig. S2A). Pressure data were converted to millibar and linearly interpolated to a 15-min resolution using the R package *zoo*. Surface solar radiation downward was also converted to PPFD (μmol m^{-2} s^{-1}) and interpolated to a 15-min resolution using the function *calc_light_merged* of the R package *streamMetabolizer* (Supporting Information Fig. S2B, S2C; Appling, Read, et al. 2018). We estimated the theoretical saturation concentrations of O_2 (O_{2sat}) from atmospheric pressure and water temperature data using the Garcia-Benson model implemented in the *streamMetabolizer* function *calc_DO_sat*.

Time steps were transformed from local time to UTC, and any days with gaps in any of the variables were removed from the dataset prior to metabolism estimation (Supporting Information Fig. S2B, S2C). All the different variables were merged into a single file; so, at the end of the data preparation process, we had 106 files (one per site) with the seven variables needed for the estimation of river ecosystem metabolism: solar time, dissolved oxygen concentration, water temperature, stage, discharge, PPFD, and saturation concentration of oxygen (Supporting Information Fig. S2B).

19395590, 2025, 5, Downloaded from https://aslopubs.onlinelibrary.wiley.com/doi/10.1002/lno.70019 by Universidad De Cantabria, Wiley Online Library on [04/07/2025]. 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

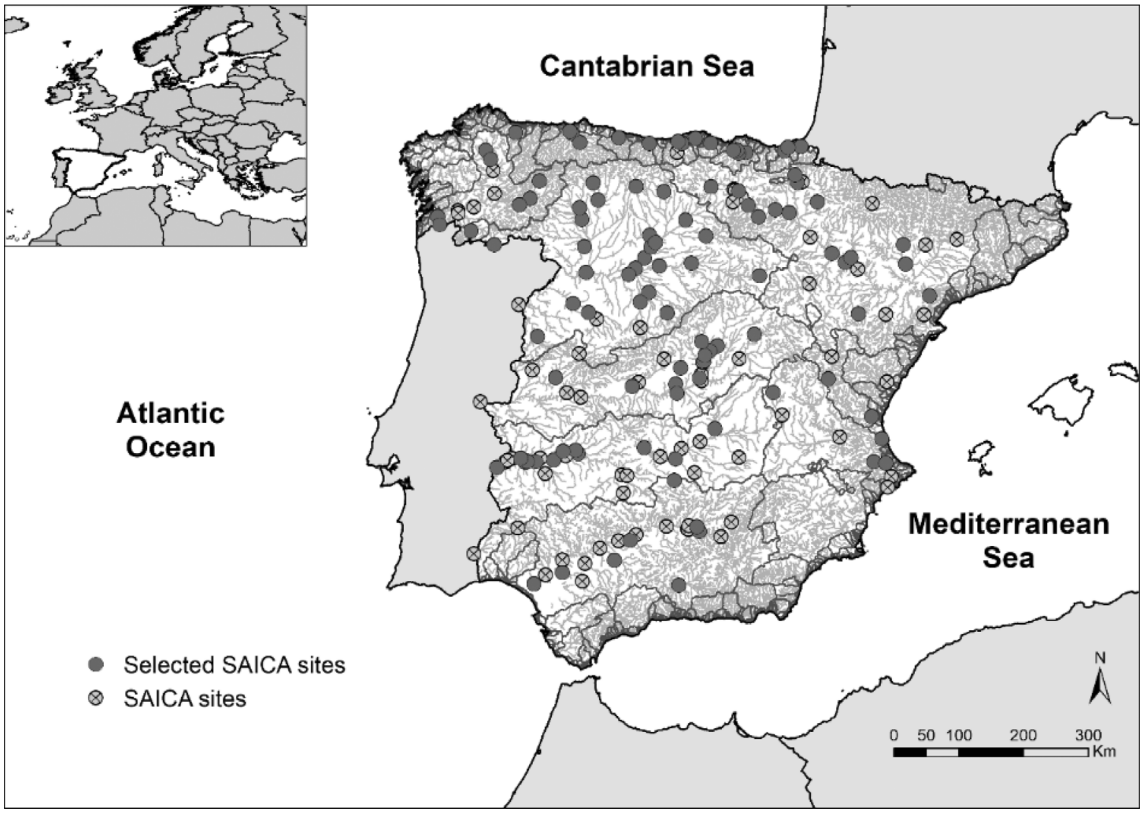

Fig. 2. Location of Spanish Automatic System of Water Quality Information (SAICA) stations in Spain and the SAICA stations selected for this study (solid gray dots).

Metabolism modeling from dissolved oxygen time series and accompanying variables

River ecosystem metabolism estimation at each site was based on the single-station open channel method (Odum 1956), using an inverse Bayesian model with Markov chain Monte Carlo implemented in the *streamMetabolizer* package to fit three unknown parameters (daily GPP, ER, and K_{600}; Hall and Hotchkiss 2017) to the observed dissolved oxygen concentrations. For a detailed description of the methodology used for daily metabolism estimation, see Supporting Information Heading Two. Metabolism was estimated over 24-h windows starting at 4:00 a.m.–3:59 a.m. the following day. Models were run in parallel on the Neptuno computing cluster at IHCantabria using up to 100 cores at the same time.

Model assessment and post-processing of river metabolism estimates

Model fitting was assessed using different metrics (Supporting Information Fig. S2F). First, MCMC chain convergence was assessed using the Gelman-Rubin parameter ($\hat{R}$) of the standard deviation of the model process error and the standard deviation of K_{600} deviations from the $K_{600} \sim Q$ relationship (Appling, Read, et al. 2018) and only models with $\hat{R} < 1.2$ in both these parameters were kept. Moreover, the relationship between ER and K_{600} was assessed, and models with high correlation between these two parameters were discarded (Spearman correlation > 0.7) as it indicates poorly fitted estimates of both (Appling, Hall, et al. 2018; Appling, Read, et al. 2018). The variability of K_{600} was also used to estimate model performance, as it is a physically constrained parameter that should not vary much. Models with a difference between the 5^{th} and 95^{th} quantiles of $K_{600} > 50$ were also discarded. After model quality assessment, 103 sites were kept for further analysis, with a mean of 1259 d of metabolism estimation per site (min. 238–max. 3516), across a mean of 6.4 different years (2–13). Results of GPP and ER with the wrong sign (GPP < −0.5 and ER > 0.5) were deleted, while values lower than those thresholds were replaced with zeros. Days with $R^2 < 0.7$ (for the observed dissolved oxygen relative to the modeled dissolved oxygen, as calculated by the *streamMetabolizer* model) were removed, as well as days when the integrated reach distance was too large (> 50 km) to yield reliable metabolic estimates (for integrated distance estimation see Supporting Information Heading Three). After that, a site with less than 1 week of data was discarded (Supporting Information Fig. S2D).

After applying these quality control criteria, 102 sites were kept for further analysis. The mean number of valid days of river ecosystem metabolism estimation per site was 966 (18–2988), across a mean of 6.22 different years (1–13) over the

19395590, 2025, 5, Downloaded from https://aslopubs.onlinelibrary.wiley.com/doi/10.1002/lno.70019 by Universidad De Cantabria, Wiley Online Library on [04/07/2025]. 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

period 2000–2020, with only five sites having less than 6 months of daily river metabolism data (Supporting Information Fig. S3). Metabolism estimates were transformed to carbon estimates using the atomic and molecular weights of C and O_2, respectively, and the photosynthetic (1.2) and respiratory (0.85) coefficients (Bott 2007).

Identifying the main environmental controls of river metabolism

To identify the relationship between the dynamic variables, static factors, and anthropogenic impacts, we selected a set of variables that describe the regional characteristics as well as the short-term reach dynamics. Land uses were downloaded (National Geographic Institute 2020) and estimated along the river catchment using a spatially explicit Synthetic River Network constructed from a Digital Elevation Map at 25 m resolution and the software NetMap (Benda et al. 2007). Topographic and anthropogenic variables (catchment area, distance to the sea, hydrological alteration, etc.) were also derived from this synthetic network. We included explicit spatial and temporal variables (longitude, latitude, and day of the year) to account for the possible effect of other factors not included in this study. Site and year were not added to the models as Random Forest cannot extrapolate outside the boundaries of the calibration dataset, thus limiting the potential for extrapolation to other reaches. Daily GPP was added to the ER model as an explanatory variable, as autotrophic respiration may represent a significant portion of total ecosystem respiration (Griffiths et al. 2013; Demars et al. 2015; Arroita et al. 2019).

Outliers were identified and removed from the dynamic variables using boxplots, and pairwise collinearity between variables was assessed, removing variables with Spearman correlation > 0.7 and variance inflation factor > 5. Three predictive variables were discarded following these criteria (elevation, atmospheric pressure, and day length). The selected predictors (Supporting Information Table S1) included a set of dynamic variables at a daily time scale, which characterize hydrology (stage, discharge), light, and water quality (temperature, conductivity, turbidity, pH). The rest of the variables were divided into two groups: static regional characteristics, which included geography and topography; and anthropogenic variables, describing three types of human-related impacts (hydrological alteration, land use, and water pollution).

Additionally, to study the effect of short- and long-term environmental variability on daily river ecosystem metabolism, we calculated the mean of the dynamic variables over five different periods (0, 5, 10, 20, and 40 d) and used them as explanatory variables in different models, following the work of Rodríguez-Castillo 2017. We used the means instead of the maximum or minimum value of the period to better integrate multiday legacy effects. The different lags were selected attending to the available data and the duration of different processes, from short-term disturbances (5 and 10 d) to longer-term growth and senescence dynamics (20–40 d).

Random forest (RF) was selected to explore the main drivers of variability of GPP and ER variability due to its high power and flexibility in analyzing large datasets. It is a supervised non-parametric machine learning algorithm that combines different classification or regression trees, especially indicated to find relationships in very large datasets with numerous explanatory variables of different types (static, dynamic, continuous, proportions, etc.), accepting non-normal variables and effectively detecting non-linear and partial relationships (Peñas et al. 2014; Rodríguez-Castillo et al. 2019; Segatto et al. 2021). Random forest combines the results of numerous decision trees developed using a random subset of the data with replacement. Each tree is created by sequentially splitting the data space by the best explanatory variable at different sequential nodes until reaching a terminal node (leaf) with the highest data purity. Although RF models are flexible, they are still sensitive to variable autocorrelation and heteroscedasticity (Breiman et al. 2018; Segatto et al. 2021).

To avoid spatiotemporal biases, only river metabolism data between 2004 and 2016 were selected as this period contained the highest number of study sites. To reduce temporal autocorrelation, random training datasets of 7500 metabolism observations were selected for model calibration. Another 2500 were selected for model prediction and overfitting testing (Diamond et al. 2021) while heteroscedasticity was reduced by applying a logarithmic transformation ($\log_{10}$) to GPP and ER. Overfitting was controlled by using an ensemble of 500 trees (individual models) developed using a random set of data points with replacement and a random one-third of the explanatory variables available for data splitting at each tree node.

Five different RF models were fitted for GPP and ER, one using the dynamic variables of the same day (Lag 0) while the other four models were calibrated with the mean value of the different sets of days explained above (Lags 5, 10, 20, 40). Five hundred regression trees were generated for each RF model by drawing a random subsample of the training dataset. One-third of the explanatory variables were selected at each node split. We started with an RF including all explanatory variables, and we sequentially simplified it by removing the least important explanatory variable (the one with the least node purity on the model) and retraining the model until we reached the model presenting the lowest root mean square error (RMSE) on the test dataset. Variable importance was assessed using the mean increase in node purity (i.e., how well the variable splits the data at a node), measured by the residual sum of squares. Additionally, the spatiotemporal dependence of dynamic variables was analyzed using RF, with the objective of determining the relative importance of temporal and spatial gradients in river ecosystem metabolism (Supporting Information Heading Four). All models were performed using the R package *randomForest* (Breiman et al. 2018).

19395590, 2025, 5, Downloaded from https://aslopubs.onlinelibrary.wiley.com/doi/10.1002/lno.70019 by Universidad De Cantabria, Wiley Online Library on [04/07/2025]. 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

Results

River ecosystem metabolism showed high variability among sites. Mean daily GPP across all sites was 1.02 ± 0.76 g C m^{-2} d^{-1}; the average per site ranged from 0.01 ± 0.01 to 3.61 ± 1.26 g C m^{-2} d^{-1}. By comparison, mean daily ER across all sites was -2.30 ± 1.32 g C m^{-2} d^{-1}; the average per site ranged from -0.04 ± 0.02 to -5.39 ± 1.80 g C m^{-2} d^{-1}. Net ecosystem production ranged from -4.26 ± 1.84 to 1.46 ± 1.64 g C m^{-2} d^{-1} with a mean of -1.27 ± 1.12 g C m^{-2} d^{-1} across all sites. Most sites (94%) were heterotrophic, that is, net ecosystem production < 0 (Supporting Information Fig. S4C). High GPP (i.e., > 4 g C m^{-2} d^{-1}) was more common on northern sites in comparison to central and southern catchments (Supporting Information Fig. S4C, S6D), while ER seemed more randomly distributed across the study region (Supporting Information Fig. S4B). Due to the higher GPP, sites in the north of Spain were less heterotrophic than in the rest of the area (Supporting Information Fig. S4C).

The Lag 40 GPP model had the highest explained variance and smallest RMSE (Table 1), although the difference in RMSE among the models was small and the additional explained variance in the Lag 40 compared to Lag 0 models was 9.5%. This increase was reflected in a slight improvement in the fit between observations and predictions, reducing the underestimation of GPP daily rates (the slopes of the linear models increased from 0.74 to 0.82), especially at their upper range (Supporting Information Fig. S8A, S8C). Likewise, the Lag 40 model explained the largest ER variance with the smallest RMSE. Here, the percentage of variance explained by the ER models steadily increased from 61.09% to 82.28% (a 34.7% increase) with the integration of longer periods, surpassing the variance explained by the Lag 40 GPP model (Table 1). This increase in explained variance was reflected in a large decrease in the prediction error across the whole range of observations (Supporting Information Fig. S8B, S8D), although the Lag 40 model still slightly underestimated the ER rates. The number of variables retained by the simplified model as well as their order of importance were similar in all the RF models fitted for GPP and ER (Table 1).

The Lag 40 simplified GPP model retained more explanatory variables than the ER one (16 vs. 11) but explained less variance. The main predictor for GPP was mean stage, closely followed by distance to the sea and light. Agricultural land use in a 200 m buffer at each bank and day of the year were the fourth and fifth most important drivers (Supporting Information Fig. S5A). Variables related to water quality (conductivity, pH, turbidity) were also important, as well as the proportion of human land uses (agriculture and urban lands). For ER, stage and GPP were the most important drivers followed by water conductivity and pH. Static variables such as distance to the sea and latitude were important but less relevant than the first three variables. Turbidity, water temperature, and discharge were also important, while anthropogenic factors were less important predictors of ER (Supporting Information Fig. S5B).

Mean hydrology (stage and discharge) of the previous 40 d positively influenced GPP until it reached a peak at a stage of ~ 1 m and discharge of ~ 20 m^3 s^{-1}. At higher discharges and stages, GPP decreased (Fig. 3a, h). On the other hand, distance to the sea negatively affected GPP, with an abrupt decrease in GPP in sites > 200 km from the sea (Fig. 3b). The relationship between GPP and light was positive, with a saturation point at high light intensities (> 500 μmol m^{-2} s^{-1}; Fig. 3c). Gross primary production was also highly influenced by the day of the year, with productivity rising in the spring and peaking in the summer, then declining in the autumn and reaching a low point at the beginning of winter (Fig. 3e). The relationship of GPP with pH was logistical, with an abrupt increase around pH 7 and reaching saturation at pH 8 (Fig. 3f). Gross primary production increased abruptly with conductivity, following an inverse exponential curve until it reached a saturation point around 1000 μS cm^{-1} (Fig. 3g). Gross primary production decreased with higher turbidity values (Fig. 3i), while water temperature was a positive driver of GPP, following a logistical relationship with a plateau in productivity at 20–25°C (Fig. 3j). Additionally, GPP was also impacted by agricultural land uses, decreasing abruptly when the percentage of agricultural lands surrounding the river reach exceeded 25% (Fig. 3d). GPP also decreased with the percentage of urban

Table 1. Number of variables retained in the model, percentage of variance explained, and root mean square error (RMSE) on the test dataset of the five random forest (RF) models created for gross primary production (GPP) and ecosystem respiration (ER). The models with the smallest RMSE are presented in bold.

	GPP			ER		
RF model	*N* variables	% Var. expl	RMSE	*N* variables	% Var. expl	RMSE
Lag 0	20	70.39	0.28	24	61.09	0.29
Lag 5	25	68.66	0.30	15	71.49	0.22
Lag 10	22	71.18	0.29	15	75.96	0.19
Lag 20	21	73.64	0.28	21	78.53	0.18
Lag 40	**16**	**77.09**	**0.25**	**11**	**82.28**	**0.17**

19395590, 2025, 5, Downloaded from https://aslopubs.onlinelibrary.wiley.com/doi/10.1002/lno.70019 by Universidad De Cantabria, Wiley Online Library on [04/07/2025]. 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

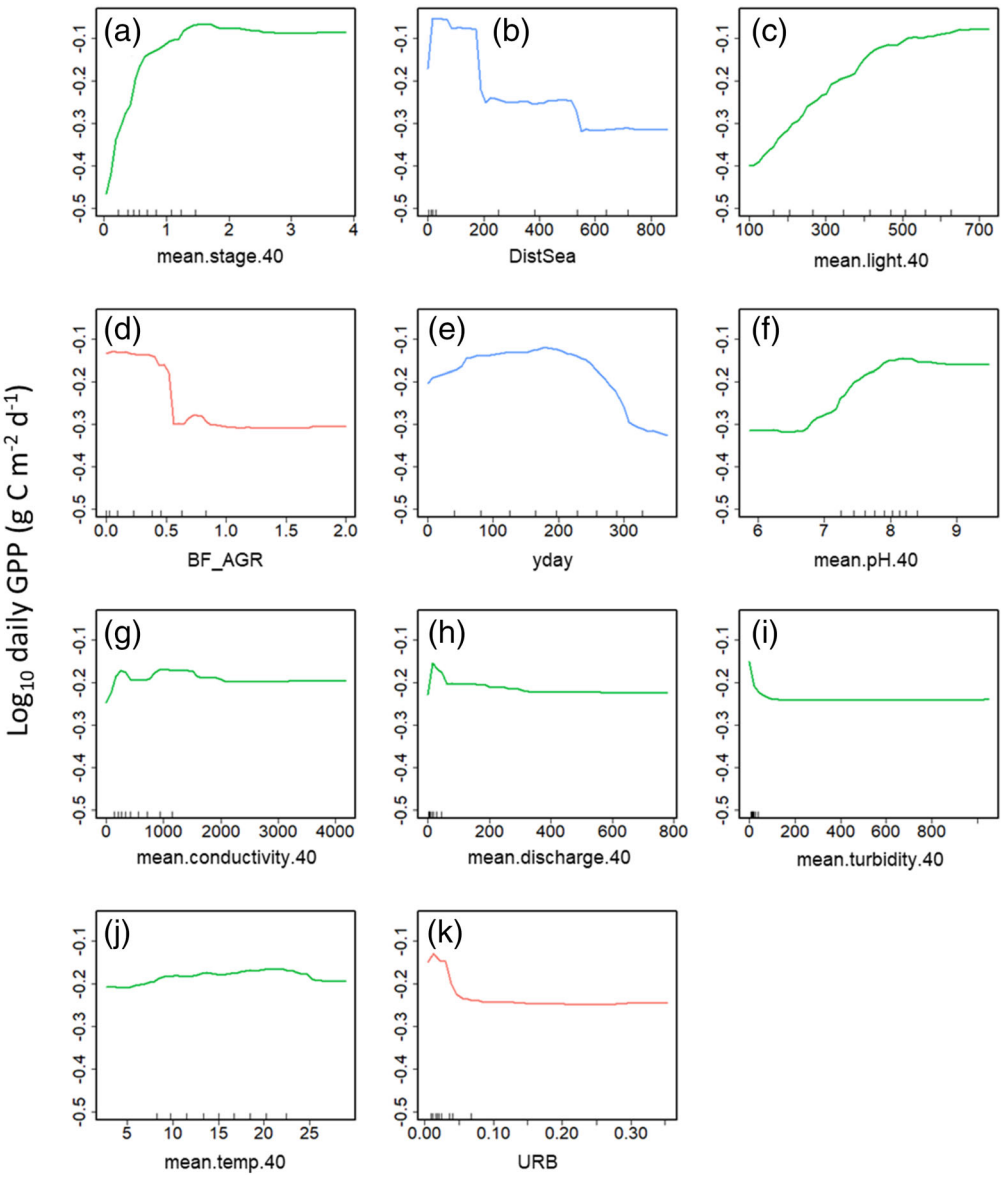

Fig. 3. Partial dependence plots of the 11 variables that explained up to 80% of the increase in node purity in the Lag 40 model for gross primary production (GPP), sorted from most important (top left corner) to least important (bottom right corner). The partial dependence plots for rest of the variables retained by the model are available in the Supporting Information Fig. S6. (**a**) A 40-d mean stage (mean.stage.40, m); (**b**) distance to the sea (DistSea, km); (**c**) 40-d mean PPFD (mean.light.40, μmol m^{-2} s^{-1}); (**d**) proportion of agricultural lands in a 200 m buffer surrounding the river reach (BF_AGR); (**e**) day of the year (yday); (**f**) 40-d mean pH (mean.pH.40); (**g**) 40-d mean conductivity (mean.conductivity.40, μS cm^{-1}); (**h**) 40-d mean discharge (mean.discharge.40, m^3 s^{-1}); (**i**) 40-d mean turbidity (mean.turbidity.40, NTU); (**j**) 40-d mean water temperature (mean.temp.40, °C); (**k**) proportion of urban lands in the upstream catchment (URB). Ticks in the *x*-axis of the plots indicate the deciles for the explanatory variable.

lands in the upstream watershed (Fig. 3k); yet other anthropogenic impacts had negligible effects despite being retained by the model (Supporting Information Fig. S6B, S6C).

Ecosystem respiration was positively related to the mean stage of the prior 40 d following an inverse exponential curve with a plateau around 1.5 m (Fig. 4a). Similarly, ER also presented a positive relationship with the mean 40-d GPP, with ER reaching a plateau at $\sim$ 4 g O_2 m^{-2} d^{-1} (1.25 g C m^{-2} d^{-1}) of GPP (Fig. 4b). ER was also positively related to conductivity, reaching a maximum at $\sim$ 1200 μS cm^{-1} (Fig. 4c). On the other hand, ER was negatively related to pH, with an abrupt decrease of respiration beyond a pH of $\sim$ 7.5 (Fig. 4d). ER rates were lower closer to the sea (< 50 km, Fig. 4e), and increased further inland, while also showing a negative relationship with latitude (Fig. 4f), with lower rates observed in the most northern catchments. The relationship between ER and discharge was also

19395590, 2025, 5, Downloaded from https://aslopubs.onlinelibrary.wiley.com/doi/10.1002/lno.70019 by Universidad De Cantabria, Wiley Online Library on [04/07/2025]. 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

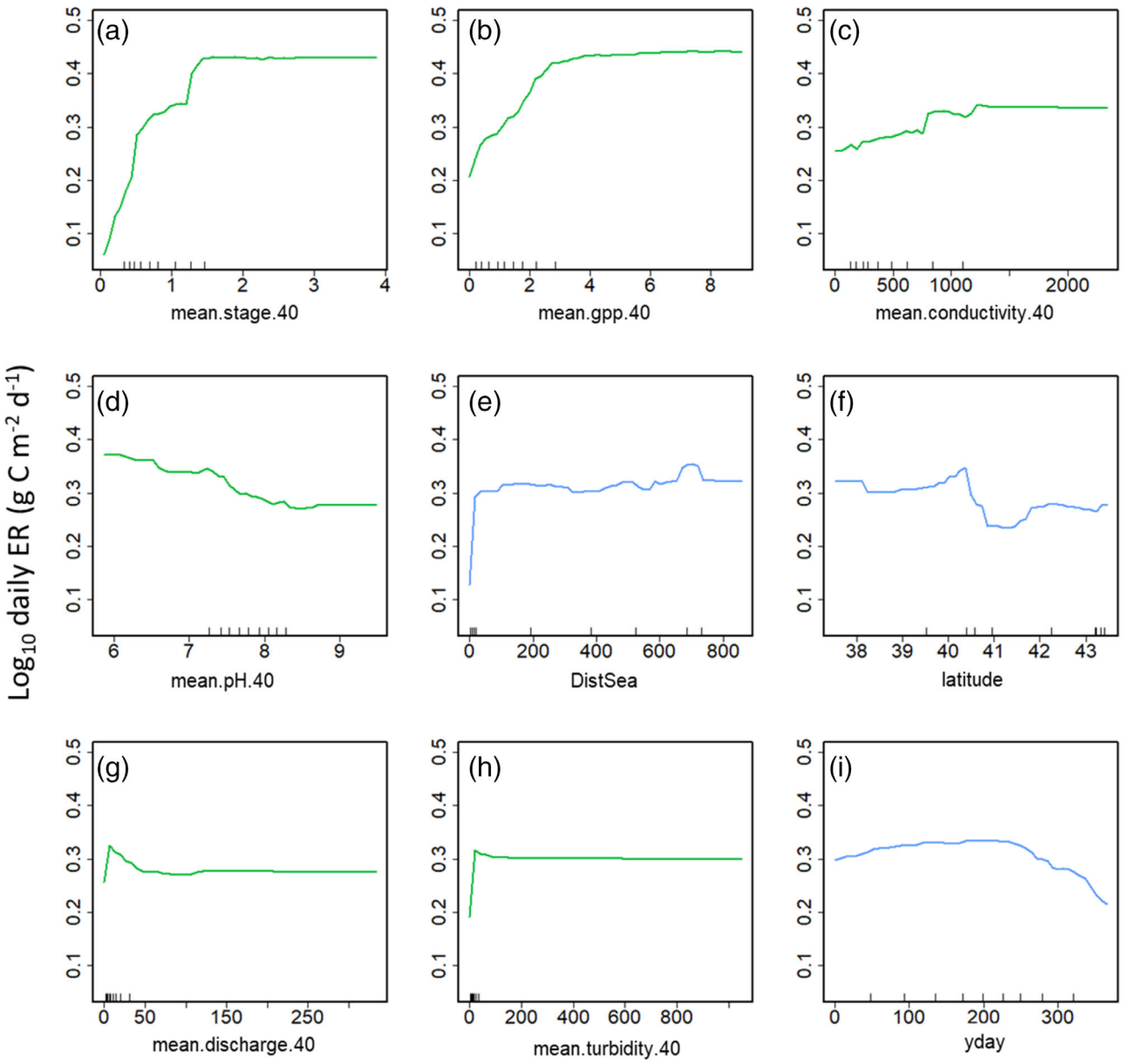

Fig. 4. Partial dependence plots of the nine variables that account for 80% of the increase in node purity in the Lag 40 model for daily ecosystem respiration (ER) rates, sorted from most important (top left corner) to least important (bottom right corner). The partial dependence plots of the other two variables retained by the model are available in the Supporting Information Fig. S7. (**a**) 40-d mean stage (mean.stage.40, m); (**b**) 40-d mean gross primary production (mean.gpp.40, g O_2 m^{-2} d^{-1}); (**c**) 40-d mean conductivity (mean.conductivity.40, μS cm^{-1}); (**d**) 40-d mean pH (mean.pH.40); (**e**) distance to the sea (DistSea, km); (**f**) latitude; (**g**) 40-d mean discharge (mean.discharge.40, m^3 s^{-1}); (**h**) 40-d mean turbidity (mean.turbidity.40, NTU); (**i**) day of the year (yday). Ticks in the *x*-axis of the plots indicate the deciles for the variable.

positive, with ER peaking at ~ 20 m^3 s^{-1} and declining thereafter (Fig. 4g). Contrarily to GPP, ER was positively related to turbidity (Fig. 4h). ER increased steadily through spring, reaching a peak in summer and then abruptly decreased during autumn (Fig. 4i). ER was not strongly influenced by temperature, although rates increased sharply between 6°C and 7°C, before leveling off at temperatures beyond 10°C (Supporting Information Fig. S7A). Land uses and human impacts did not have a substantial effect in the ER model, with the proportion of broadleaf forests in the upstream river basin being the only variable kept by the model (Supporting Information Fig. S7B).

Discussion

Our study is the first attempt to model river ecosystem metabolism at large spatiotemporal gradients in the Iberian

19395590, 2025, 5, Downloaded from https://aslopubs.onlinelibrary.wiley.com/doi/10.1002/lno.70019 by Universidad De Cantabria, Wiley Online Library on [04/07/2025]. 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

Peninsula, using environmental variables collected by automatic monitoring stations and remote sensing resources. The generated models can predict metabolism rates consistently across a wide range of environmental conditions. This predictive ability is highly valuable for large spatiotemporal studies that require interpolation of incomplete time series, extrapolation of river metabolism in space and time, or global change predictions. Our results indicate that RF is a good tool for the study of river ecosystem metabolism drivers and patterns in large environmental gradients. However, as RF models are not able to extrapolate outside the range of the input variables, it is necessary to develop large datasets including very diverse environmental conditions to ensure accurate river metabolism predictions.

Overall, our findings showed that dynamic variables were more important for river ecosystem metabolism control, but static variables (distance to the sea and day of the year) were also highly influential drivers of GPP. Further, average environmental conditions over longer periods showed stronger relationships to metabolic rates than same-day conditions, particularly for ER. Such delays suggest that metabolic rates are also linked to longer-term dynamics of biomass accrual, senescence, and community succession. Yet, there are also limitations to consider in our analysis, a main one being the use of stage as a proxy for depth in metabolism calculation, which may result in an under- or overestimation of metabolic rates. We also lack information about the methodology used for stage-discharge calibration at each SAICA site, adding uncertainty to our results. Furthermore, due to lack of data, our analysis did not include important drivers of metabolic rates such as nutrients, dissolved organic matter, or biomass estimations. Finally, despite the high predictive power of our Lag 40 models, high metabolic rates were still underestimated, although this bias may be reduced by using higher predictor integration times (lags > 40 d). Nevertheless, our approach produced reasonable metabolism predictions at large spatiotemporal scales and thus seems promising for application to other regions with a different set of environmental gradients.

Dynamic variables

As reflected in our conceptual framework and previous studies, river metabolism across the Iberian Peninsula was mainly controlled by dynamic variables, with GPP primarily related to light availability and stage (Young and Huryn 1996, 1999; Mulholland et al. 2001; McTammany et al. 2007). The high relative importance of stage may be an artifact of metabolism estimation, as this is one of the main variables used in the Bayesian model. Our results suggested that GPP reached an ecosystem-level saturation point at a daily mean PPFD of $\sim 500\ \mu mol\ m^{-2}\ s^{-1}$. This leveling off of GPP may be determined by many different factors, such as self-shading, environmental stress, composition and abundance of the autotrophic community, or even photoinhibition (Acuña et al. 2004; Townsend et al. 2018; Boix Canadell et al. 2021). The relationship between GPP and light availability was also mediated by day of the year, as light availability was highly seasonally dependent (Supporting Information Heading Four, Table S2, Fig. S9A, S11A; Savoy et al. 2019; Savoy and Harvey 2021).

The relationship of GPP and ER with stage was also positive, reaching a maximum and then stabilizing. Because high discharge events that caused biofilm scouring were eliminated prior to analysis, this result may imply that GPP is limited by the capacity of light to penetrate the water column (Kirk et al. 2021), which changes spatially and temporally (Supporting Information Heading Four, Figs. S10B, S10C, S13A, S13B) with river size and discharge due to turbidity (Uehlinger et al. 2003; Blaszczak et al. 2019; Chowanski et al. 2020). This turbidity could be caused by high sediment and organic matter loads, or by the increase of algal concentration in the water column, producing a density-dependent control on GPP. On the other hand, turbidity was positively related to ER at the lower end of the range, perhaps because turbidity is also related to higher nutrient and organic matter availability, which might boost ER (Young and Huryn 1996; Palmer and Ruhi 2019). Discharge was also an important variable in daily river metabolism dynamics (Figs. 3h, 4g), behaving similarly to water stage for GPP and ER, which is consistent as both stage and discharge change with river size and precipitation regimes (Uehlinger et al. 2003; Acuña et al. 2004).

After stage, the second most important driver of ER was GPP. Although the relationship between GPP and ER may be due to their dependence on the same environmental variables, this may also suggest that autotrophic organisms are a major source of respiration (Alnoee et al. 2015) or that they provide organic matter in the biofilm for heterotrophic consumption. Disentangling the contribution of both autotrophic and heterotrophic respiration to ER is a difficult task and beyond the objectives of this study. Further research is needed to fully understand the relationship between autotrophic and heterotrophic community dynamics and river ecosystem metabolism (Munn et al. 2020; Segatto et al. 2020). Regardless, other recent studies in large rivers indicate a close coupling between GPP and ER (Roley et al. 2023), and our findings suggest that this may similarly be true for rivers across the Iberian Peninsula.

While less important than light and stage, variation in water chemistry also influenced metabolic patterns across sites. For example, GPP tended to increase with pH, while ER showed the opposite relationship (Figs. 3f, 4d). However, the observed variation in pH may be a result of metabolic rates, as ER produces CO_2, lowering pH, while photosynthesis consumes CO_2, consequently raising pH (Caldeira and Wickett 2003). Both lower pH and higher ER may also be related to higher dissolved organic carbon, although we did not have the data to test this hypothesis. Despite pH being, at least in part, a result of metabolic activity, it is an important

19395590, 2025, 5, Downloaded from https://aslopubs.onlinelibrary.wiley.com/doi/10.1002/lno.70019 by Universidad De Cantabria, Wiley Online Library on [04/07/2025]. 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

predictor in both GPP and ER models, and thus, its inclusion may improve model predictions. Rates of GPP and ER also increased across sites with conductivity. This relationship exhibited a prominent spatial dependency (Supporting Information Heading Four, Table S2, Figs. S10A, S12) as conductivity depends on catchment geology and increases with catchment size. These findings collectively suggest that larger rivers tend to accumulate more dissolved ions and nutrients from their surrounding terrestrial ecosystems, and these inputs boost rates of metabolic activity.

Water temperature exhibited a positive relationship with both GPP and ER, with a saturation point at $\sim 20^{\circ}C$ and $\sim 12^{\circ}C$, respectively. ER increased abruptly around $\sim 5^{\circ}C$ while GPP increased more gradually (Fig. 3j and Supporting Information Fig. S7A). The different water temperature sensitivities of GPP and ER may help explain their different seasonal patterns, with ER being less sensitive to colder temperatures than GPP, which could have implications for river carbon cycling under changing climate conditions (Song et al. 2018). Due to the abrupt increase of ER at cold temperatures, a potential increase in temperature may skew the GPP/ER balance toward respiration, thus transitioning metabolic activity toward higher heterotrophy in a large fraction of the river network length (i.e., headwaters and small tributaries; Song et al. 2018). Water temperature variability in the Iberian Peninsula, as well as light, was highly seasonally dependent, increasing toward summer and decreasing in winter (Supporting Information Heading Four, Figs. S9B, S11B). Thus, the relationship observed between metabolic rates and temperature was mainly caused by seasonal cycles and not spatial environmental variability.

Static variables: Regional characteristics and anthropogenic impacts

The third most important factor explaining GPP variability was distance to the sea, such that GPP reached its peak closer to the river mouth, with an abrupt decrease at distances larger than 200 km (Fig. 3b). These patterns are consistent with other studies in Spain reporting increases in GPP downstream (Rodríguez-Castillo et al. 2019). More broadly, such patterns are consistent with the river continuum concept (Vannote et al. 1980), which predicts that GPP should increase in a downstream direction; yet this general pattern can vary in response to other factors, such as land use in the catchment (Rodríguez-Castillo et al. 2019). Moreover, in the case of the Iberian Peninsula, this pattern could also reflect a geographical bias as reaches in smaller Cantabrian river basins (northern Spain) are located more closely to the sea and are generally very productive.

River metabolism was not substantially affected by human activities, although our study may not have captured significant undisturbed–impacted gradients as SAICA sites are located in large rivers typically affected by a variety of anthropogenic impacts. Anthropogenic impacts may also impact metabolic rates in an indirect way, by affecting local variables, such as light, water temperature, or discharge, which our study approach was not suited to detect. The most important anthropogenic impacts for river metabolism were land uses, mainly agriculture. The proportion of agricultural land surrounding the river reach had an important negative effect on GPP (Fig. 3d), contrary to the results of previous studies. Other studies conclude that the loss of riparian forest together with the increased nutrient loads due to agriculture practices may result in higher GPP rates (Bernot et al. 2010; Griffiths et al. 2013; Alnoee et al. 2015). However, opposite patterns may appear due to interactions among several drivers altered by agricultural activity, including elevated nutrient concentrations, which can boost metabolic rates (Aristi et al. 2016) but also inputs of herbicides and pesticides (Kosinski 1984; Morin et al. 2010; Aristi et al. 2016) and/or increases in fine sediments and turbidity, all of which may reduce GPP and ER (Atkinson et al. 2008; Matthaei et al. 2010). These contradictory results may reflect an incomplete understanding of the effect of agriculture on river metabolism (Finlay 2011; Bernhardt et al. 2018).

Integration of temporal environmental variability

Our models indicated that GPP temporal variability was better captured by light and stage variation integrated over longer time periods, as shown by the lowest RMSE in the Lag 40 RF model. The Lag 40 model was less prone to underestimate higher GPP rates (> 4 g C m^{-2} d^{-1}; Supporting Information Fig. S8A, S8C), suggesting higher GPP rates were the result of longer-term biomass growth dynamics, which were better captured by integrating environmental conditions over longer periods. However, the increase in explained variance between the worst (Lag 0) and best model (Lag 40) was only 9.5%, which may indicate that a mixture of short-term and long-term variables is driving daily GPP rates (e.g., daily light availability and autotroph growth dynamics). A previous study also found a relationship between GPP and light over the previous week (Blaszczak et al. 2019) mediated by biomass growth. The small improvement from the Lag 0 to the Lag 40 model also suggests GPP was mainly controlled by simple organisms (e.g., unicellular algae) with short life cycles, high growth rates, and quick recovery times after perturbations (Uehlinger 2006; Acuña et al. 2015).

Likewise, ER variability was best captured by variables integrating larger time periods, with the Lag 40 RF model being the one with the lowest RMSE, improving the explained variance of the worst model (Lag 0) by almost 35%. These findings indicate that daily ER rates are likely controlled by a more complex set of longer-term environmental drivers than GPP, increasing respiration resistance to short-term fluctuations (Rodríguez-Castillo 2017). The longer response period may indicate that ER is more dependent on the seasonal deposition of organic matter, coupled with the delayed response of biomass growth and decay of a diverse array of simple organisms

19395590, 2025, 5, Downloaded from https://aslopubs.onlinelibrary.wiley.com/doi/10.1002/lno.70019 by Universidad De Cantabria, Wiley Online Library on [04/07/2025]. 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

(bacteria, algae, etc.), presenting different growth rates and sensitivity to perturbation (Munn et al. 2020; Bertuzzo et al. 2022). Despite its high predictive power, the Lag 40 model still consistently underestimated higher ER rates (< -5 g C m^{-2} d^{-1}), suggesting that higher biomass growth integration periods (> 40 d) may be needed to estimate these respiration peaks accurately. Nevertheless, this result is a promising avenue of new research on the temporal variability of river metabolic rates in relation to river ecosystem resilience and resistance to changing environmental conditions. However, our data had gaps and did not allow us to analyze larger time frames (60 – 365 d) and describe the effects of seasonal environmental patterns, which should be a next step in our understanding of legacy effects on metabolism.

Conclusions

This study presents a modeling approach that can more accurately predict daily river metabolism across large environmental gradients by integrating potential predictors over longer periods. Improving metabolism prediction accuracy, even by small amounts, is essential to reduce the prediction error of metabolic rates at regional or global scales. Here, the models that included longer environmental legacies considerably improved the prediction accuracy and were less prone to underestimating metabolic rates, highlighting the integrative nature of river ecosystem metabolism. Long-term biomass dynamics should be further explored to unravel their relative contributions to riverine metabolism rates.

Moreover, this study shows for the first time the key driving factors of river metabolism in larger Iberian rivers (stream order ≥ 5). As expected, dynamic variables were generally the most important predictors of river ecosystem metabolism, although distance to the sea also shaped spatial patterns along river networks. Therefore, daily metabolic rates depended on a mix of dynamic variables integrated over weeks, regulated by spatial and temporal patterns such as position in the river network and season. These spatiotemporal patterns may help describe general metabolism regimes in entire river networks and find common drivers at regional and even global scales.

Our study also suggests that RF models may be effective tools for predicting river metabolism across large environmental gradients. Our overall approach requires two main elements: a deep understanding of river ecosystem metabolism patterns and drivers, and large databases with continuous oxygen records from rivers spanning gradients in size and environmental condition, for which monitoring networks are essential. Despite the large environmental dataset used, this work still presents some sampling biases, as it is skewed toward small and mid-sized rivers (when put in a global context) in a temperate–Mediterranean climatic gradient affected by diverse anthropogenic impacts (water abstraction, agriculture, etc.). Further study on the geographical and temporal complexity of metabolic factors in other river types is needed to extrapolate general trends if we want to characterize river ecosystem metabolism patterns at a global scale.

Author Contributions

Amaia A. Rodeles: Conceptualization; investigation; formal analysis; methodology; writing—original draft preparation; writing—review and editing; visualization. Francisco J. Peñas: Conceptualization; methodology; writing—review and editing; funding acquisition; project administration. Maite Arroita: methodology; writing—review and editing. José Barquín: Conceptualization; funding acquisition; supervision; methodology; writing—review and editing.

Acknowledgments

The authors acknowledge the financial support received from the Government of Cantabria through the Fénix Programme. This publication is also part of the I+D+i project RIFFLE PID2020-114427RJ-I00 funded by MCIN/AEI/10.13039/501100011033. José Barquín received funding from the I+D+i project BIORESP PID2023-150641OB-I00 funded by MICIU/AEI/10.13039/501100011033 and ERDF, EU.

Conflicts of Interest

None declared.

References

Acuña, V., M. Casellas, N. Corcoll, X. Timoner, and S. Sabater. 2015. "Increasing Extent of Periods of No Flow in Intermittent Waterways Promotes Heterotrophy." *Freshwater Biology* 60: 1810–1823. https://doi.org/10.1111/fwb.12612.

Acuña, V., A. Giorgi, I. Muñoz, U. Uehlinger, and S. Sabater. 2004. "Flow Extremes and Benthic Organic Matter Shape the Metabolism of a Headwater Mediterranean Stream." *Freshwater Biology* 49: 960–971. https://doi.org/10.1111/j.1365-2427.2004.01239.x.

Allen, A. P., J. F. Gillooly, and J. H. Brown. 2005. "Linking the Global Carbon Cycle to Individual Metabolism." *Functional Ecology* 19: 202–213. https://doi.org/10.1111/j.1365-2435.2005.00952.x.

Alnoee, A. B., T. Riis, M. R. Andersen, A. Baattrup-Pedersen, and K. Sand-Jensen. 2015. "Whole-Stream Metabolism in Nutrient-Poor Calcareous Streams on Öland, Sweden." *Aquatic Sciences* 77: 207–219. https://doi.org/10.1007/s00027-014-0380-5.

Appling, A. P., R. O. Hall, C. B. Yackulic, and M. Arroita. 2018. "Overcoming Equifinality: Leveraging Long Time Series for Stream Metabolism Estimation." *Journal of Geophysical Research: Biogeosciences* 123: 624–645. https://doi.org/10.1002/2017JG004140.

Appling, A. P., J. S. Read, and L. A. Winslow. 2018. "Data Descriptor: The Metabolic Regimes of 356 Rivers in the

19395590, 2025, 5, Downloaded from https://aslopubs.onlinelibrary.wiley.com/doi/10.1002/lno.70019 by Universidad De Cantabria, Wiley Online Library on [04/07/2025]. 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

United States." *Scientific Data* 5: 1–14. https://doi.org/10.1038/sdata.2018.292.

Aristi, I., M. Casellas, A. Elosegi, et al. 2016. "Nutrients Versus Emerging Contaminants—or a Dynamic Match between Subsidy and Stress Effects on Stream Biofilms." *Environmental Pollution* 212: 208–215. https://doi.org/10.1016/j.envpol.2016.01.067.

Arroita, M., A. Elosegi, and R. O. Hall. 2019. "Twenty Years of Daily Metabolism Show Riverine Recovery Following Sewage Abatement." *Limnology and Oceanography* 64: S77–S92. https://doi.org/10.1002/lno.11053.

Arroita, M., L. Flores, and A. Larrañaga. 2017. "Water Abstraction Impacts Stream Ecosystem Functioning Via Wetted-Channel Contraction." *Freshwater Biology* 62: 243–257. https://doi.org/10.1111/fwb.12864.

Atkinson, B. L., M. R. Grace, B. T. Hart, and K. E. N. Vanderkruk. 2008. "Sediment Instability Affects the Rate and Location of Primary Production and Respiration in a Sand-Bed Stream." *Journal of the North American Benthological Society* 27: 581–592. https://doi.org/10.1899/07-143.1.

Bai, Y.,J. Cao, E. Chen, et al. 2020. "Estimate of Vegetation Production of Terrestrial Ecosystem." In Advanced Remote Sensing, Terrestrial Information Extraction and Applications, 2nd edn., edited by S. Liang and J. Wang, 581–620. London, UK: Academic Press.

Battin, T. J., Lauerwald, R., Bernhardt, E. S., and others. 2023. River Ecosystem Metabolism and Carbon Biogeochemistry in a Changing World. *Nature* 613: 449–459 . https://doi.org/10.1038/s41586-022-05500-8

Benda, L., D. Miller, K. Andras, P. Bigelow, G. Reeves, and D. Michael. 2007. "NetMap: A New Tool in Support of Watershed Science and Resource Management." *Forest Science* 53: 206–219. https://doi.org/10.1093/forestscience/53.2.206

Bernhardt, E., P. Savoy, and M. Vlah. 2022. "Light and Flow Regimes Regulate the Metabolism of Rivers." *Proceedings of the National Academy of Sciences of the United States of America* 119, no. 8: e2121976119. https://doi.org/10.1073/pnas.2121976119/-/DCSupplemental.

Bernhardt, E. S., J. B. Heffernan, and N. B. Grimm. 2018. "The Metabolic Regimes of Flowing Waters." *Limnology and Oceanography* 63: S99–S118. https://doi.org/10.1002/lno.10726.

Bernot, M. J., D. J. Sobota, and R. O. Hall. 2010. "Inter-Regional Comparison of Land-Use Effects on Stream Metabolism." *Freshwater Biology* 55: 1874–1890. https://doi.org/10.1111/j.1365-2427.2010.02422.x.

Bertuzzo, E., E. R. Hotchkiss, and A. Argerich. 2022. "Respiration Regimes in Rivers: Partitioning Source-Specific Respiration from Metabolism Time Series." *Limnology and Oceanography* 67, no. 11: 1–15. https://doi.org/10.1002/lno.12207.

Blaszczak, J. R., J. M. Delesantro, D. L. Urban, M. W. Doyle, and E. S. Bernhardt. 2019. "Scoured or Suffocated: Urban Stream Ecosystems Oscillate between Hydrologic and Dissolved Oxygen Extremes." *Limnology and Oceanography* 64: 877–894. https://doi.org/10.1002/lno.11081.

Blaszczak, J. R., C. B. Yackulic, R. K. Shriver, and R. O. Hall. 2023. "Models of Underlying Autotrophic Biomass Dynamics Fit to Daily River Ecosystem Productivity Estimates Improve Understanding of Ecosystem Disturbance and Resilience." *Ecology Letters* 26, no. 9: 1–13. https://doi.org/10.1111/ele.14269.

Boix Canadell, M., L. Gómez-Gener, A. J. Ulseth, M. Clémençon, S. N. Lane, and T. J. Battin. 2021. "Regimes of Primary Production and Their Drivers in Alpine Streams." *Freshwater Biology* 66: 1449–1463. https://doi.org/10.1111/fwb.13730.

Boscolo-Galazzo, F., K. A. Crichton, S. Barker, and P. N. Pearson. 2018. "Temperature Dependency of Metabolic Rates in the Upper Ocean: A Positive Feedback to Global Climate Change?" *Global and Planetary Change* 170: 201–212. https://doi.org/10.1016/j.gloplacha.2018.08.017.

Bott, T. L. 2007. "Primary Productivity and Community Respiration." In Methods in Stream Ecology, 2nd edn., edited by F. R. Hauer and G. A. Lamberti, 663–690. London, UK: Academic Press.

Breiman, L., A. Cutler, A. Liaw, and M. Wiener. 2018. "Breiman and Cutler's Random Forests for Classification and Regression." 29. https://cran.r-project.org/web/packages/randomForest/randomForest.pdf

Buitenhuis, E. T., T. Hashioka, and C. Le Quéré. 2013. "Combined Constraints on Global Ocean Primary Production Using Observations and Models." *Global Biogeochemical Cycles* 27: 847–858. https://doi.org/10.1002/gbc.20074.

Caldeira, K., and M. E. Wickett. 2003. "Anthropogenic Carbon and Ocean pH." *Nature* 425: 365. https://doi.org/10.1038/425365a

Center for Study and Experimentation of Construction Works. 2020. Anuario de aforos. Accessed September 9, 2020. https://ceh.cedex.es/anuarioaforos/default.asp.

Chowanski, K., L. Kunza, G. Hoffman, L. Genzoli, and E. Stickney. 2020. "River Management Alters Ecosystem Metabolism in a Large Oligotrophic River." *Freshwater Science* 39: 534–548. https://doi.org/10.1086/710082.

Cole, J. J., Y. T. Prairie, and N. F. Caraco. 2007. "Plumbing the Global Carbon Cycle: Integrating Inland Waters into the Terrestrial Carbon Budget." *Ecosystems* 10: 171–184. https://doi.org/10.1007/s10021-006-9013-8.

Demars, B. O. L., J. Thompson, and J. R. Manson. 2015. "Stream Metabolism and the Open Diel Oxygen Method: Principles, Practice, and Perspectives." *Limnology and Oceanography: Methods* 13: 356–374. https://doi.org/10.1002/lom3.10030.

Diamond, J. S., F. Moatar, M. J. Cohen, et al. 2021. "Metabolic Regime Shifts and Ecosystem State Changes Are Decoupled in a Large River." *Limnology and Oceanography* 67, no. S1: 1–17. https://doi.org/10.1002/lno.11789.

19395590, 2025, 5, Downloaded from https://aslopubs.onlinelibrary.wiley.com/doi/10.1002/lno.70019 by Universidad De Cantabria, Wiley Online Library on [04/07/2025]. 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

Field, C. B., M. J. Behrenfeld, J. T. Randerson, and P. Falkowski. 1998. "Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components." *Science* 281: 237–240. https://doi.org/10.1126/science.281.5374.237.

Finlay, J. C. 2011. "Stream Size and Human Influences on Ecosystem Production in River Networks." *Ecosphere* 2, no. 8: art87. https://doi.org/10.1890/es11-00071.1.

Fuß, T., B. Behounek, A. J. Ulseth, and G. A. Singer. 2017. "Land Use Controls Stream Ecosystem Metabolism by Shifting Dissolved Organic Matter and Nutrient Regimes." *Freshwater Biology* 62: 582–599. https://doi.org/10.1111/fwb.12887.

Genzoli, L., and R. O. Hall. 2016. "Shifts in Klamath River Metabolism Following a Reservoir Cyanobacterial Bloom." *Freshwater Science* 35: 795–809. https://doi.org/10.1086/687752.

Griffiths, N. A., J. L. Tank, T. V. Royer, et al. 2013. "Agricultural Land Use Alters the Seasonality and Magnitude of Stream Metabolism." *Limnology and Oceanography* 58: 1513–1529. https://doi.org/10.4319/lo.2013.58.4.1513.

Hall, R. O., and E. R. Hotchkiss. 2017. Stream Metabolism. In Methods in Stream Ecology Volume 2, edited by G. A. Lamberti and F. R. Hauer, 219–233, Academic Press. London. Third Edition.

Izagirre, O., U. Agirre, M. Bermejo, J. Pozo, and A. Elosegi. 2008. "Environmental Controls of Whole-Stream Metabolism Identified from Continuous Monitoring of Basque Streams." *Journal of the North American Benthological Society* 27: 252–268. https://doi.org/10.1899/07-022.1.

Kirk, L., R. T. Hensley, P. Savoy, J. B. Heffernan, and M. J. Cohen. 2021. "Estimating Benthic Light Regimes Improves Predictions of Primary Production and Constrains Light-Use Efficiency in Streams and Rivers." *Ecosystems* 24: 825–839. https://doi.org/10.1007/s10021-020-00552-1.

Koenig, L. E., A. M. Helton, P. Savoy, et al. 2019. "Emergent Productivity Regimes of River Networks." *Limnology and Oceanography Letters* 4: 173–181. https://doi.org/10.1002/lol2.10115.

Kosinski, R. J. 1984. "The Effect of Terrestrial Herbicides on the Community Structure of Stream Periphyton." *Environmental Pollution Series A, Ecological and Biological* 36: 165–189. https://doi.org/10.1016/0143-1471(84)90097-7.

Leopold, L. B., and T. Maddock, Jr. 1953. The Hydraulic Geometry of Stream Channels and Some Physiographic Implications. Washington, DC.: U.S. Government Printing Office.

Matthaei, C. D., J. J. Piggott, and C. R. Townsend. 2010. "Multiple Stressors in Agricultural Streams: Interactions among Sediment Addition, Nutrient Enrichment and Water Abstraction." *Journal of Applied Ecology* 47: 639–649. https://doi.org/10.1111/j.1365-2664.2010.01809.x.

McTammany, M. E., E. F. Benfield, and J. R. Webster. 2007. "Recovery of Stream Ecosystem Metabolism from Historical Agriculture." *Journal of the North American Benthological Society* 26: 532–545. https://doi.org/10.1899/06-092.1.

Ministry for the Ecological Transition and the Demographic Challenge. 2020. Red de alterta (SAICA). Accessed June 19, 2020. https://www.miteco.gob.es/es/agua/temas/estado-y-calidad-de-las-aguas/aguas-superficiales/programas-seguimiento/saica.aspx.

Morin, S., S. Pesce, A. Tlili, M. Coste, and B. Montuelle. 2010. "Recovery Potential of Periphytic Communities in a River Impacted by a Vineyard Watershed." *Ecological Indicators* 10: 419–426. https://doi.org/10.1016/j.ecolind.2009.07.008.

Mulholland, P. J., C. S. Fellows, and J. L. Tank. 2001. "Inter-Biome Comparison of Factors Controlling Stream Metabolism." *Freshwater Biology* 46: 1503–1517. https://doi.org/10.1046/j.1365-2427.2001.00773.x.

Munn, M. D., R. W. Sheibley, I. R. Waite, and M. R. Meador. 2020. "Understanding the Relationship between Stream Metabolism and Biological Assemblages." *Freshwater Science* 39, no. 4: 680–692. https://doi.org/10.1086/711690.

Muñoz Sabater, J. 2019. "ERA5-Land Hourly Data from 1950 to Present." Copernicus Climate Change Service. Climate Data Store. https://doi.org/10.24381/cds.e2161bac.

National Geographic Institute. 2020. Sistema de Información de Ocupación del Suelo de España (SIOSE). Accessed September 9, 2020. https://www.siose.es/web/guest/base-de-datos.

Odum, H. T. 1956. "Primary Production in Flowing Waters." *Limnology and Oceanography* 1: 102–117. https://doi.org/10.4319/lo.1956.1.2.0102.

Palmer, M., and A. Ruhi. 2019. "Linkages between Flow Regime, Biota, and Ecosystem Processes: Implications for River Restoration." *Science* 365: 1–13. https://doi.org/10.1126/science.aaw2087.

Peñas, F. J., and J. Barquín. 2019. "Assessment of Large-Scale Patterns of Hydrological Alteration Caused by Dams." *Journal of Hydrology* 572: 706–718. https://doi.org/10.1016/j.jhydrol.2019.03.056.

Peñas, F. J., J. Barquín, T. H. Snelder, D. J. Booker, and C. Álvarez. 2014. "The Influence of Methodological Procedures on Hydrological Classification Performance." *Hydrology and Earth System Sciences Discussions* 11: 945–985. https://doi.org/10.5194/hessd-11-945-2014.

Potter, S., T. Randerson, B. Field, A. Matson, and H. A. Mooney. 1993. "Terrestrial Ecosystem Production: A Process Model Based on Global Satellite and Surface Data." *Global Biogeochemical Cycles* 7: 811–841. https://doi.org/10.1029/93GB02725

Raymond, P. A., J. Hartmann, and R. Lauerwald. 2013. "Global Carbon Dioxide Emissions from Inland Waters." *Nature* 503: 355–359. https://doi.org/10.1038/nature12760.

Rivas-Martínez, S. 1987. Memoria del mapa de series de vegetación de España, edited by Instituto para la Conservación

19395590, 2025, 5, Downloaded from https://aslopubs.onlinelibrary.wiley.com/doi/10.1002/lno.70019 by Universidad De Cantabria, Wiley Online Library on [04/07/2025]. 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

de la Naturaleza, 1st ed. Ministerio de Agricultura, Pesca y Alimentación.

Roberts, B. J., P. J. Mulholland, and W. R. Hill. 2007. "Multiple Scales of Temporal Variability in Ecosystem Metabolism Rates: Results from 2 Years of Continuous Monitoring in a Forested Headwater Stream." *Ecosystems* 10: 588–606. https://doi.org/10.1007/s10021-007-9059-2.

Rodríguez-Castillo, T. 2017. Determining Ecosystem Functioning Spatio-Temporal Patterns in Atlantic Rivers. University of Cantabria.

Rodríguez-Castillo, T., E. Estévez, A. M. González-Ferreras, and J. Barquín. 2019. "Estimating Ecosystem Metabolism to Entire River Networks." *Ecosystems* 22: 892–911. https://doi.org/10.1007/s10021-018-0311-8.

Roley, S. S., R. O. Hall, W. Perkins, V. A. Garayburu-Caruso, and J. C. Stegen. 2023. "Coupled Primary Production and Respiration in a Large River Contrasts with Smaller Rivers and Streams." *Limnology and Oceanography* 68, no. 11: 2461–2475. https://doi.org/10.1002/lno.12435.

Savoy, P., A. P. Appling, J. B. Heffernan, et al. 2019. "Metabolic Rhythms in Flowing Waters: An Approach for Classifying River Productivity Regimes." *Limnology and Oceanography* 64: 1835–1851. https://doi.org/10.1002/lno.11154.

Savoy, P., and J. W. Harvey. 2021. "Predicting Light Regime Controls on Primary Productivity across CONUS River Networks." *Geophysical Research Letters* 48: 1–10 https://doi.org/10.1029/2020GL092149.

Segatto, P. L., T. J. Battin, and E. Bertuzzo. 2020. "Modeling the Coupled Dynamics of Stream Metabolism and Microbial Biomass." *Limnology and Oceanography* 65: 1573–1593. https://doi.org/10.1002/lno.11407.

Segatto, P. L., T. J. Battin, and E. Bertuzzo. 2021. "The Metabolic Regimes at the Scale of an Entire Stream Network Unveiled through Sensor Data and Machine Learning." *Ecosystems* 24: 1792–1809. https://doi.org/10.1007/s10021-021-00618-8.

Song, C., W. K. Dodds, and J. Rüegg. 2018. "Continental-Scale Decrease in Net Primary Productivity in Streams Due to Climate Warming." *Nature Geoscience* 11: 415–420. https://doi.org/10.1038/s41561-018-0125-5.

Townsend, S. A., I. T. Webster, M. A. Burford, and J. Schult. 2018. "Effects of Autotrophic Biomass and Composition on Photosynthesis, Respiration and Light Utilisation Efficiency for a Tropical Savanna River." *Marine and Freshwater Research* 69: 1279–1289. https://doi.org/10.1071/MF17172.

Uehlinger, U. 2006. "Annual Cycle and Inter-Annual Variability of Gross Primary Production and Ecosystem Respiration in a Floodprone River during a 15-Year Period." *Freshwater Biology* 51: 938–950. https://doi.org/10.1111/j.1365-2427.2006.01551.x.

Uehlinger, U., B. Kawecka, and C. T. Robinson. 2003. "Effects of Experimental Floods on Periphyton and Stream Metabolism below a High Dam in the Swiss Alps (River Spöl)." *Aquatic Sciences* 65: 199–209. https://doi.org/10.1007/s00027-003-0664-7.

Vannote, R., G. W. Minshall, K. W. Cummins, J. R. Sedell, and C. E. Cushing. 1980. "The River Continuum Concept." *Canadian Journal of Fisheries and Aquatic Sciences* 37: 130–137. https://doi.org/10.1139/f80-017

Young, R. G., and A. D. Huryn. 1996. "Interannual Variation in Discharge Controls Ecosystem Metabolism along a Grassland River Continuum." *Canadian Journal of Fisheries and Aquatic Sciences* 53: 2199–2211. https://doi.org/10.1139/f96-186.

Young, R. G., and A. D. Huryn. 1999. "Effects of Land Use on Stream Metabolism and Organic Matter Turnover." *Ecological Applications* 9: 1359–1376. https://doi.org/10.1890/1051-0761(1999)009[1359:EOLUOS]2.0.CO;2.

Zeileis, A., G. Grothendieck, J. A. Ryan, J. M. Ulrich, and F. Andrews. 2021. "S3 Infrastructure for Regular and Irregular Time Series (Z's Ordered Observations)." 1–75. https://cran.r-project.org/web//packages/zoo/zoo.pdf

Supporting Information

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

Submitted 10 January 2024
Revised 27 June 2024
Accepted 16 February 2025

19395590, 2025, 5, Downloaded from https://aslopubs.onlinelibrary.wiley.com/doi/10.1002/lno.70019 by Universidad De Cantabria, Wiley Online Library on [04/07/2025]. 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