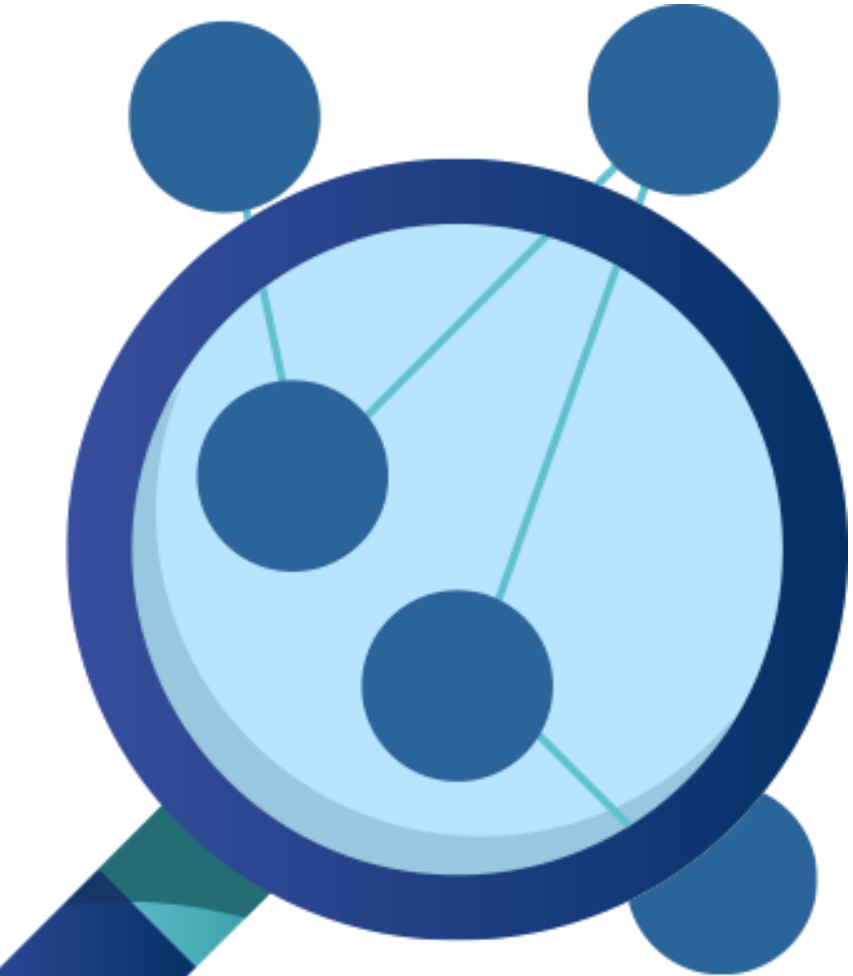
Modelling workflow to forecast short-term changes in dissolved organic matter in source waters

Deliverable 5.1

WP5 SOURCE - Adaptation options and source protection



Authors: Ricardo Paíz, Daniel Mercado-Bettín, Rafael Marcé, Eleanor Jennings, Valerie McCarthy

Date: October 2025

Funded by the European Union

GRANT AGREEMENT NUMBER	101081728		
FULL TITLE / ACRONYM	INNOVATIVE TOOLS TO CONTROL ORGANIC MATTER AND DISINFECTION BYPRODUCTS IN DRINKING WATER / intoDBP		
START DATE	01/12/2022	**DURATION**	48 months
END DATE	30/11/2026		
PROJECT URL	**https://intodbp.eu/**		
DELIVERABLE	**D5.1**		
WORK PACKAGE	**WP5**		
CONTRACTUAL DATE OF DELIVERY	**30-11-2025**		
ACTUAL DATE OF DELIVERY	**06-05-2026**		
NATURE	**REPORT**	**DISSEMINATION LEVEL**	**PUBLIC**
LEAD BENEFICIARY	**DCU**		
RESPONSIBLE AUTHOR	**Ricardo Paíz**		
CONTRIBUTIONS FROM	**ICRA**		
ABSTRACT	This deliverable introduces two complementary workflows for modelling and forecasting dissolved organic matter (DOM) in source waters (lake/reservoirs and rivers, respectively) using different modelling approaches (machine learning and process-based models) and a set of different proxies for DOM. The frameworks were applied across multiple study sites representing contrasting hydroclimatic and anthropogenic conditions, allowing assessment of the environmental drivers and variability that shape DOM dynamics in systems where rising organic matter can affect drinking water supply. Short-term daily forecasts and seasonal-scale predictions were generated and evaluated, demonstrating the potential of these approaches to anticipate near-term changes in DOM levels. These modelling workflows can produce practical and actionable information for water managers and catchment regulators seeking to enhance operational readiness, strengthen adaptive strategies, and improve resilience to fluctuations in DOM water quality.		

Disclaimer

Any dissemination of results reflects only the author's view and the European Commission is not responsible for any use that may be made of the information it contains.

© intoDBP Consortium, 2022
This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation, or both. Reproduction is authorised provided the source is acknowledged.

TABLE OF CONTENTS

LIST OF FIGURES *5*

LIST OF TABLES *7*

LIST OF ABBREVIATIONS *8*

EXECUTIVE SUMMARY *12*

1. INTRODUCTION *13*

1.1 Background 13
1.2 Objectives 14

2. METHODS *16*

2.1 Study sites 16
2.2 Data availability on DOM 19
2.3 Seasonal modelling and forecasting workflow 21
2.4 Daily-scale modelling and forecasting workflow 26

3. RESULTS *30*

3.1 Seasonal-scale simulations, insights and predictions 30
3.2 Daily-scale simulations, insights and predictions 43

4. DISCUSSION *52*

4.1 Scientific Insights 52
4.2 Accuracy of the workflows 53
4.3 Scalability 54
4.4 Limitations and future research 55

5. CONCLUSIONS *56*

REFERENCES *57*

LIST OF FIGURES

Figure 1. Lough Feeagh and Sau Reservoir, two contrasting freshwater ecosystems including their catchment extents and topographical gradient. 16
Figure 2. Lough Feeagh and Sau Reservoir, two contrasting freshwater ecosystems including their catchment extents and land cover. The land covers were based on the CORINE 2018 dataset. 17
Figure 3. The Glenamong River including its river network and topographical gradient. The catchment outlet and river gauge location are indicated in the figure. 18
Figure 4. Workflow implemented for simulating fDOM in both lake study sites. 23
Figure 5. Selecting influential drivers to predict fDOM. 24 drivers from various sources were used to train the four ML models that directly provide feature importance: Random Forest (RF), eXtreme Gradient Boosting (XGBoost), Light Gradient Boosting (LGB), CatBoost (CTB). These included 8 climate variables, 8 soil variables, 2 outputs from hydrologic and water quality modeling, 5 outputs from lake modeling, and the cosine of the Julian day. The most relevant drivers were identified based on their contribution to node purity or gain contribution, with only those exceeding a 5% variable importance being selected. For both case studies, the key drivers were soil temperature at 100–255 cm, soil moisture at 28–100 cm and 100–255 cm, and Julian day. Additionally, lake surface water temperature, inflow DOC concentration, and soil temperature at 28–100 cm were selected for Feeagh, while lake volume was selected for Sau. 30
Figure 6. Driver influence in predicting fDOM. It presents the partial dependence plots for (a.) Lough Feeagh and (b.) Sau Reservoir based on the random forest (RF) model, and SHAP beeswarm plots for (c.) Lough Feeagh and (d.) Sau Reservoir based on the CatBoost (CTB) model. In Feeagh, soil temperature at the deepest layer strongly influenced fDOM predictions, while in Sau, soil moisture at the deepest layer plays a significant role due to water availability constraints. In both cases, increases in these key drivers correspond to increases in fDOM. In both study sites, Julian day was a relevant factor, although the influence of seasonality on fDOM predictions was more evident in Feeagh. 32
Figure 7. fDOM predictions obtained from the CatBoost model using different driver sets. Here the training (blue, 85% of the time series) and testing (violet or green, 15% of the time series) periods are shown. The training and testing periods are completely independent and follow the hold-out period method, and model performance metrics (R^2, KGE, and RMSE) are calculated for both periods. a. Training and testing results for Feeagh: using the 8 most influential drivers (violet) and using only easily accessible and reanalysis data (soil temperature, soil moisture and Julian day) from those influential drivers (green). 34
Figure 8. fDOM predictions obtained from the CatBoost model using different driver sets. Here the training (blue, 85% of the time series) and testing (violet or green, 15% of the time series) periods are shown. The training and testing periods are completely independent and follow the hold-out period method, and model performance metrics (R^2, KGE, and RMSE) are calculated for both periods. b. Training and testing results for Sau: using the 5 most influential drivers (violet) and using only easily accessible and reanalysis data (soil temperature, soil moisture and Julian day) from those influential drivers (green). 35
Figure 9. Verification of the seasonal forecast product for 2-m air temperature using CRPSS for the periods 1993–2016 (upper panel) and 2017–2023 (lower panel) in Sau Reservoir catchment.

CRPSS values above zero indicate added skill relative to climatology. Values are shown as a function of initial forecast month (x-axis) and forecast lead time in months (y-axis). 36
Figure 10. Verification of the seasonal forecast product for total precipitation using CRPSS for 1993–2016 (upper panel) and 2017–2023 (lower panel) in Sau Reservoir catchment. CRPSS values above zero indicate that the seasonal forecast product outperforms climatological averages. Results are displayed across initial forecast months (x-axis) and forecast lead times from one to seven months (y-axis). 37
Figure 11. Verification of the seasonal forecast product for soil moisture (Layer 1) using CRPSS for 1993–2016 (upper panel) and 2017–2023 (lower panel) in Sau Reservoir catchment. Positive CRPSS values indicate improved performance relative to climatology. The heatmaps show forecast skill across initial forecast months (x-axis) and forecast lead times (y-axis). 38
Figure 12. Verification of the seasonal forecast product for soil temperature (Layer 1) using CRPSS for 1993–2016 (upper panel) and 2017–2023 (lower panel) in Sau Reservoir catchment. CRPSS values higher than zero indicate added value beyond climatology. Skill scores are presented for all initial forecast months (x-axis) and lead times up to seven months (y-axis). ..39
Figure 13. Verification of the seasonal forecasts of fDOM monthly average concentration predictions using CRPSS for 1993–2016 (upper panel) and 2017–2023 (lower panel) in Lough Feeagh. CRPSS values higher than zero indicate added value beyond climatology. Skill scores are presented for all initial forecast months (x-axis) and lead times up to seven months (y-axis). 41
Figure 14. Verification of the seasonal forecasts of fDOM monthly average concentration predictions using CRPSS for 1993–2016 (upper panel) and 2017–2023 (lower panel) in Sau Reservoir. CRPSS values higher than zero indicate added value beyond climatology. Skill scores are presented for all initial forecast months (x-axis) and lead times up to seven months (y-axis). 42
Figure 15. Observed and simulated daily river discharge during the model calibration and validation of the hydrology module of GWLF-DOC (a), and observed and simulated daily DOC concentration during the model calibration of the DOC module of GWLF-DOC (b) for the Glenamong catchment. 43
Figure 16. Boxplot distributions of the weather forecast product input data for GWLF-DOC (i.e., daily forecasted catchment precipitation and mean air temperature) (a and b), modelled daily catchment discharge (c) and DOC observations used for data assimilation (d) over the meteorological seasons in the forecasting period (1st March 2021 – 1st March 2023). 46
Figure 17. DOC concentration forecasts produced using the process-based model and for Site Persistence for different forecast horizons (1-, 2-, 3-,7-, 14-, 21-, 28- and 35-days ahead) throughout the entire forecasting period plotted with the observed DOC concentrations used for evaluation. The background light shade colours in the graphs indicate spring (green), summer (red), autumn (yellow) and winter (blue) seasons, respectively. Method B indicates the successful run used for the evaluation of the forecasting, and method A indicates the worst run produced during workflow development which is displayed for reference. 49
Figure 18. Root mean squared error (RMSE) and Nash-Sutcliffe coefficient (NSE) performance for the process-based workflow and for Site Persistence for DOC concentration forecasts over the forecast horizon for the meteorological seasons in the forecasting period. Method B indicates the successful run used for the evaluation of the forecasting, and method A indicates the worst run produced during workflow development which is displayed for reference. 50

Figure 19. Coefficient of determination (r^2) performance and quantified forecast uncertainty (standard deviation) of the process-based workflow andfor Site Persistence for DOC concentration forecasts over the forecast horizon for the meteorological seasons in the forecasting period. Method B indicates the successful run used for the evaluation of the forecasting, and method A indicates the worst run produced during workflow development which is displayed for reference. 51

LIST OF TABLES

Table 1. Model comparison to predict fDOM using all selected drivers in both study sites. Statistic metrics (R^2, RMSE and KGE) were calculated to compare the performance of the models during the testing (hold-out) period. The table support the selection of the best model for each study site between Random Forest (RF), eXtreme Gradient Boosting (XGBoost), Light Gradient Boosting (LGB), CatBoost (CTB), k-Nearest Neighbors (KNN), Support Vector Regression (SVR) and linear model. Overall, the best performance was found in the CatBoost model. Supplementary Table S1 contains the results during the training period. 33
Table 2. Observed and simulated daily river discharge during the model calibration and validation of the hydrology module of GWLF-DOC (a), and observed and simulated daily DOC concentration during the model calibration of the DOC module of GWLF-DOC (b) for the Glenamong catchment. 44
Table 3. Magnitudes and variability across meteorological seasons of all forcing data used to generate DOC concentration forecasts i.e., the weather forecast input data (i.e., daily forecasted total precipitation and mean air temperature), the modelled forcing hydrology (daily catchment discharge, surface runoff and baseflow), and the DOC observations used for data assimilation during the entire forecasting period. 47

LIST OF ABBREVIATIONS

Abbreviation	Full Description
AET	Actual Evapotranspiration
AWQMS	Automatic Water Quality Monitoring Station
C	Carbon
CDOM	Chromophoric Dissolved Organic Matter
CLIME	Climate and Lake Impacts in Europe
CMIP	Coupled Model Intercomparison Project
CN	Curve Number
CO_2	Carbon Dioxide
CORINE	Coordination of Information on the Environment
CRPS	Continuous Ranked Probability Score
DA	Data Assimilation
DBP	Disinfection By-product
DkIT	Dundalk Institute of Technology
DOC	Dissolved Organic Carbon
DOM	Dissolved Organic Matter

ECA&D	The European Climate Assessment & Dataset Project
EnKF	Ensemble Kalman Filter
EPA	Environmental Protection Agency
ET	Evapotranspiration
EU	European Union
FAO	Food and Agriculture Organization of the United Nations
FDOM	Fluorescent Dissolved Organic Matter
FLARE	Forecasting Lake and Reservoir Ecosystems
GCM	Global Climate Model
GHG	Greenhouse Gas
GLEON	Global Lake Ecological Observatory Network
GLM	General Lake Model
GLUE	The Generalised Likelihood Uncertainty Estimation Method
GWLF	The Generalised Watershed Loading Functions Model
GWLF-DOC	The Generalised Watershed Loading Functions Model with an Integrated DOC Module
HAA	Haloacetic Acids
HFM	High Frequency Monitoring
INCA	The Integrated Catchment Model

INCA-C	The Integrated Catchment Model with an DOC Module
IPCC	Intergovernmental Panel on Climate Change
ISIMIP	Inter-Sectoral Impact Model Intercomparison Project
ITN	Innovative Training Network
KF	Kalman Filter
NOAA	National Oceanic and Atmosphere Administration
NOAA GEFS	The National Oceanic and Atmospheric Administration Global Ensemble Forecasting System
NSE	Nash-Sutcliffe Efficiency
NYC DEP	New York City Department of Environmental Protection
OC	Organic Carbon
PET	Potential Evapotranspiration
QA	Quality Assurance
QC	Quality Control
r^2	Coefficient of Determination
RCP	Representative Concentration Pathway
RMSE	Root Mean Square Error
SA	Sensitivity Analysis
SSP	Shared Socioeconomic Pathway

SWAT	Soil and Water Assessment Tool
THMs	Trihalomethanes
TOC	Total Organic Carbon
UK	United Kingdom
USA	United States of America
WFD	Water Framework Directive
WHO	World Health Organisation

EXECUTIVE SUMMARY

Dissolved organic matter (DOM) is a key indicator of raw drinking-water quality because it affects treatability and impacts on the formation of disinfection by-products (DBPs). Its variability responds to climate, land use and hydrological conditions, which makes short-term predictions of raw water quality increasingly important. This deliverable presents two workflows that support anticipation and operational decision-making for source waters through the generation of forecasts indicative of future levels of DOM in the short-term (daily to seasonal time scales).

Two complementary modelling and forecasting workflows were developed as part of the work undertaken in WP5 of the intoDBP project. In the first workflow, we applied supervised machine learning to simulate and forecast fluorescent dissolved organic matter (fDOM) in two contrasting freshwater lake systems. This approach used environmental drivers that describe meteorological, catchment and soil processes, and was tested in a natural humic lake in Ireland and a human-influenced reservoir in Spain. The model generated seasonal forecasts at time horizons of one to seven months and can provide information on conditions when DOM is likely to increase and potentially complicate water treatment in the context of DBP formation. In the second workflow, we used a process-based hydrological and water quality model (GWLF-DOC) to predict the concentrations of dissolved organic carbon (DOC) in an Irish river. In this case, DOC served as a proxy for DOM and allowed, together with other driver datasets, the generation of daily forecasts up to 34 days ahead. These predictions offer a more operational view of near-term variability and could allow managers to detect emerging increases in organic matter and to plan daily-scale treatment responses accordingly.

Our work demonstrates that forecasts for DOM can be produced and supported by emerging technology and data availability. Moreover, these approaches can be explored in similar settings under data-limited conditions with the potential to retain skill. As part of the intoDBP project, ongoing monitoring and data acquisition in additional Irish rivers as target study sites will support broader testing of these workflows. The results presented here, however, outline a transferable pathway that can help bridge the gap between environmental modelling outputs and water management.

1. INTRODUCTION

1.1 Background

Lakes play a central role in freshwater ecosystems and global biogeochemical cycles, and they support important services such as drinking-water supply, fisheries and irrigation (Creed et al., 2018; Toming et al., 2020). Their condition, however, is increasingly influenced by climatic variability and human pressures in the surrounding catchments (Jennings et al., 2010; Meyer-Jacob et al., 2019). River systems feeding these lakes are also affected by similar pressures, and together they shape the transfer of water and organic matter from the terrestrial landscape into downstream water bodies. In this context, dissolved organic matter has emerged as a key water quality variable because it regulates light availability, thermal structure, microbial activity and a range of internal biogeochemical processes (Xenopoulos et al., 2021).

The dynamics of dissolved organic matter are driven by interactions that begin in the catchment and continue within the lake (Arvola et al., 2010; Jennings et al., 2010; Meyer-Jacob et al., 2019). Land cover, soil characteristics, hydrology and meteorological conditions regulate the mobilisation of carbon in river catchments, which then enters lakes through runoff and baseflow. Within lakes, physical and biogeochemical processes such as stratification, mixing, microbial degradation and photochemical reactions further modify the quantity and composition of dissolved organic matter (Lake et al., 2000; Anderson et al., 2023). These coupled catchment-to-lake pathways influence ecosystem stability and the suitability of water for drinking-water treatment, given that elevated levels of organic matter can increase treatment demand and the potential formation of disinfection by-products (Piai et al., 2024; Riyadh and Peleato, 2024).

Understanding and anticipating changes in dissolved organic matter at different time scales is therefore essential for effective water quality management (Bradford et al., 2020; Xu et al., 2020; Paíz et al., 2025a). However, predicting its variability remains challenging because the underlying processes are highly dynamic, non-linear and sensitive to external forcing. Rivers and lakes can respond rapidly to changes in weather, hydrology or catchment conditions, leading to short-lived but important fluctuations in organic matter. At the same time, broader seasonal patterns are shaped by climatic cycles, vegetation activity and longer-term hydrological regimes (Anderson et al., 2023).

Modelling tools offer a pathway to describe and forecast these dynamics. Process-based hydrological and lake models remain important for understanding the mechanisms that drive dissolved organic matter in both rivers and lakes, and for estimating how these systems respond to changes in climate or land use (Beven, 2009; Dietze, 2017; Carey et al., 2022). Machine learning models provide a complementary set of tools that can use large volumes of environmental drivers to predict dissolved organic matter without relying on extensive parameterisation (Harkort and Duan, 2023; Asadollah et al., 2025). These approaches provide opportunities to develop short-term and seasonal forecasts that can support proactive water management and improve the reliability of drinking-water supply.

In this deliverable, we introduce two workflows for simulating and predicting dissolved organic matter (DOM) in source waters using different proxies at two time scales, seasonal for lakes and daily for rivers. For the lake systems, we coupled process-based catchment and lake modelling with machine-learning algorithms to simulate and forecast DOM in two contrasting settings. This approach allowed us to work with a broad set of globally available drivers and to explore their influence on model behaviour, while avoiding the considerable calibration burden typically required when using process-based lake models alone. For the river site, we focused on the catchment model as a standalone tool, building on previous work that had already demonstrated its suitability for our study site. These modelling frameworks aim to provide a step forward in environmental modelling and to show how forecast outputs, at both seasonal and daily scales, can offer practical, actionable information for the management of source waters.

1.2 Objectives

The main aim of this deliverable was to develop and introduce two modelling workflows capable of simulating and forecasting DOM in source waters. These workflows were designed to capture the short-term variability of DOM using different modelling approaches and data sources, and were tested in different study sites that had measured data available to train and validate the frameworks.

Specific objectives include the following. First, to build and test a seasonal-scale workflow for simulating and predicting monthly fluorescent dissolved organic matter (fDOM) concentrations up to seven months ahead in lake systems using supervised machine

learning driven by environmental predictors from reanalysis data and process-based model outputs. This includes the identification of main drivers, the selection of the most influential drivers, model calibration through training and testing periods, and an evaluation of how accurately the workflow can predict fDOM using a reference method to contextualise performance. Second, to develop a workflow for simulating dissolved organic carbon (DOC) in rivers using a process-based hydrological and water quality model to generate daily forecasts (up to 34 days ahead). This, similarly to the first workflow, includes model calibration and validation, an assessment of predictive skill, and an examination of how these forecasts can support operational source and treatment water management tasks.

2. METHODS

2.1 Study sites

For the modelling and forecasting of DOM (through the use of fDOM) at the seasonal scale, we focused on two freshwater lake systems, one located in Ireland, Lough Feeagh, and one located in Spain, Sau Reservoir.

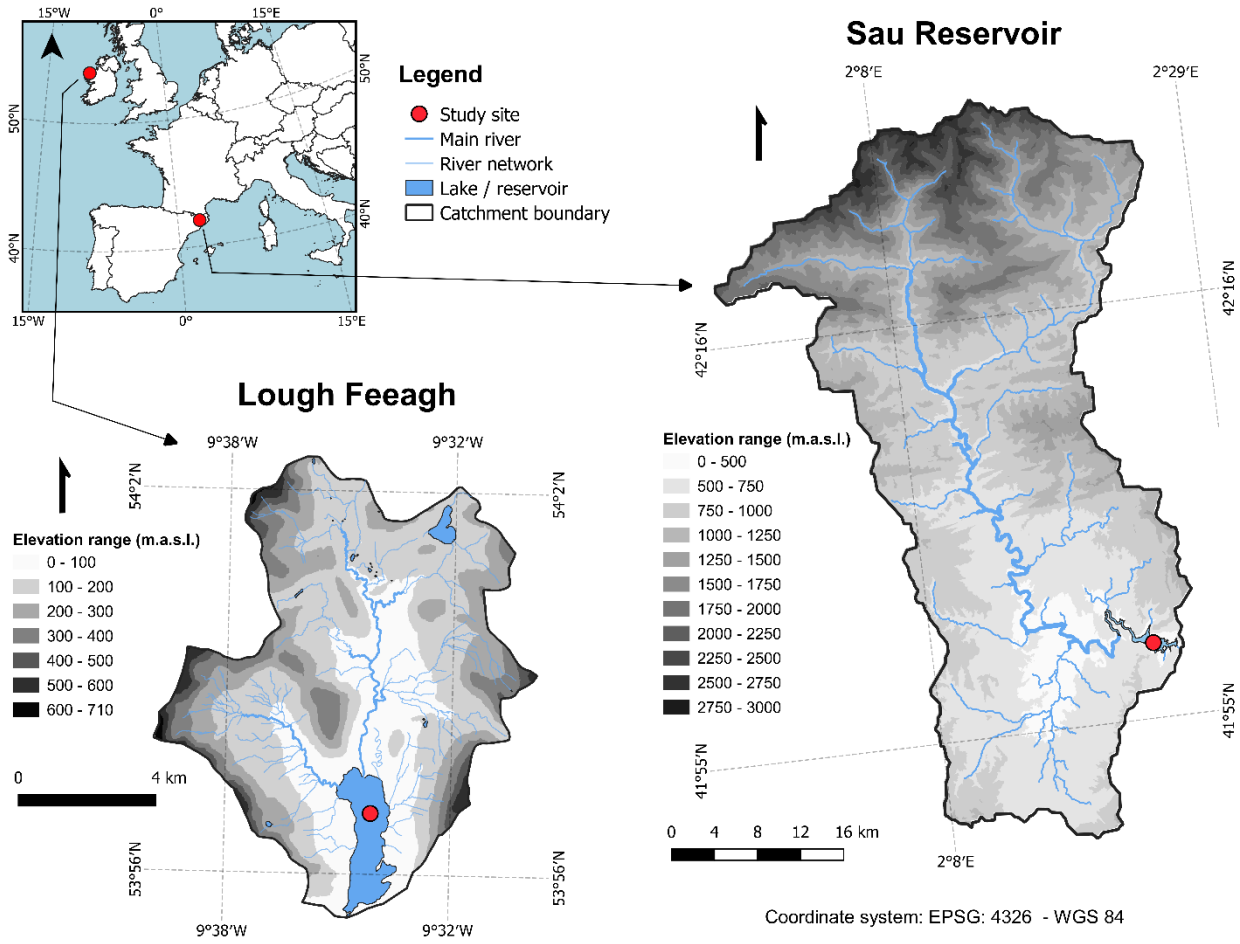


Figure 1. *Lough Feeagh and Sau Reservoir, two contrasting freshwater ecosystems including their catchment extents and topographical gradient.*

Lough Feeagh and Sau Reservoir differ in their physical and catchment attributes. Lough Feeagh is a monomictic and oligotrophic lake with a depth of 46.8 m and an area of 3.95 km², surrounded by a relatively undisturbed catchment, the Burrishoole catchment, which is dominated by peatland. The lake receives water from two main inflows, the Black and the Glenamong rivers. Sau Reservoir, in contrast, is an eutrophic and heavily regulated system with a depth of 70 m and an area of 5.8 km², and is influenced by human activities

and water abstraction. Its catchment is larger, features diverse land uses and topography, and is primarily fed by the Ter River. The Sau Reservoir is used to supply the metropolitan area of Barcelona city, and represents an excellent real-world source-to-tap case study. These sites were selected because they represent contrasting settings and therefore dissolved organic matter dynamics, providing an opportunity to test the forecasting workflows under different conditions.

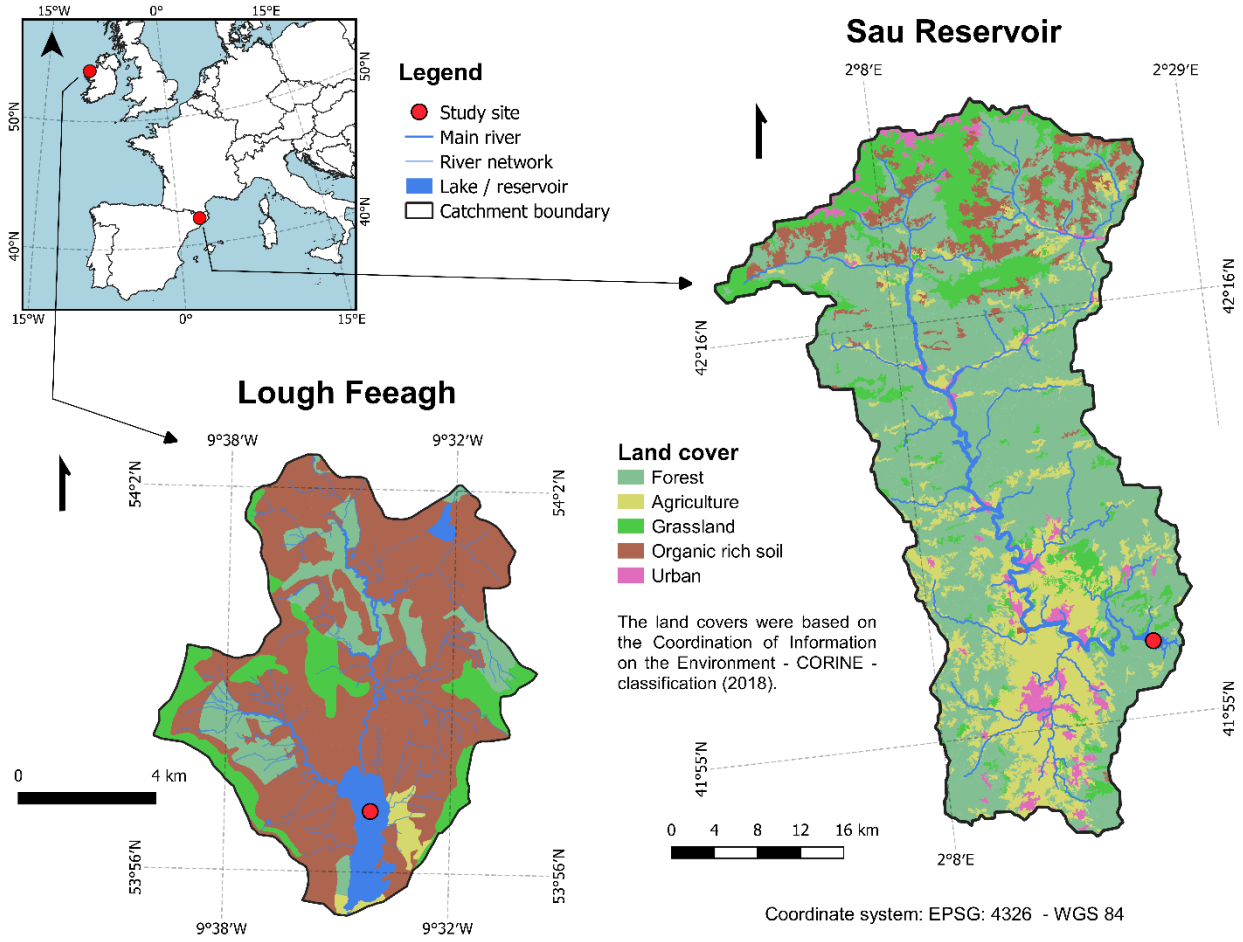


Figure 2. *Lough Feeagh and Sau Reservoir, two contrasting freshwater ecosystems including their catchment extents and land cover. The land covers were based on the CORINE 2018 dataset.*

For the modelling and forecasting of DOM (using DOC) at the daily scale, we focused on the Glenamong River in western Ireland, which is one of the two main inflows mentioned above to Lough Feeagh in the Burrishoole catchment. The river catchment is small, unpopulated and dominated by peatland and other organic soils typical of the regional landscape. Its steep terrain and climatic setting result in high variability in daily discharge and dissolved organic carbon export. Although the river is not used as a drinking water

source, it is an intensively monitored system that is representative of similar small river catchments used for drinking-water supply in the region. This made it an appropriate site for testing the daily scale modelling and forecasting workflow.

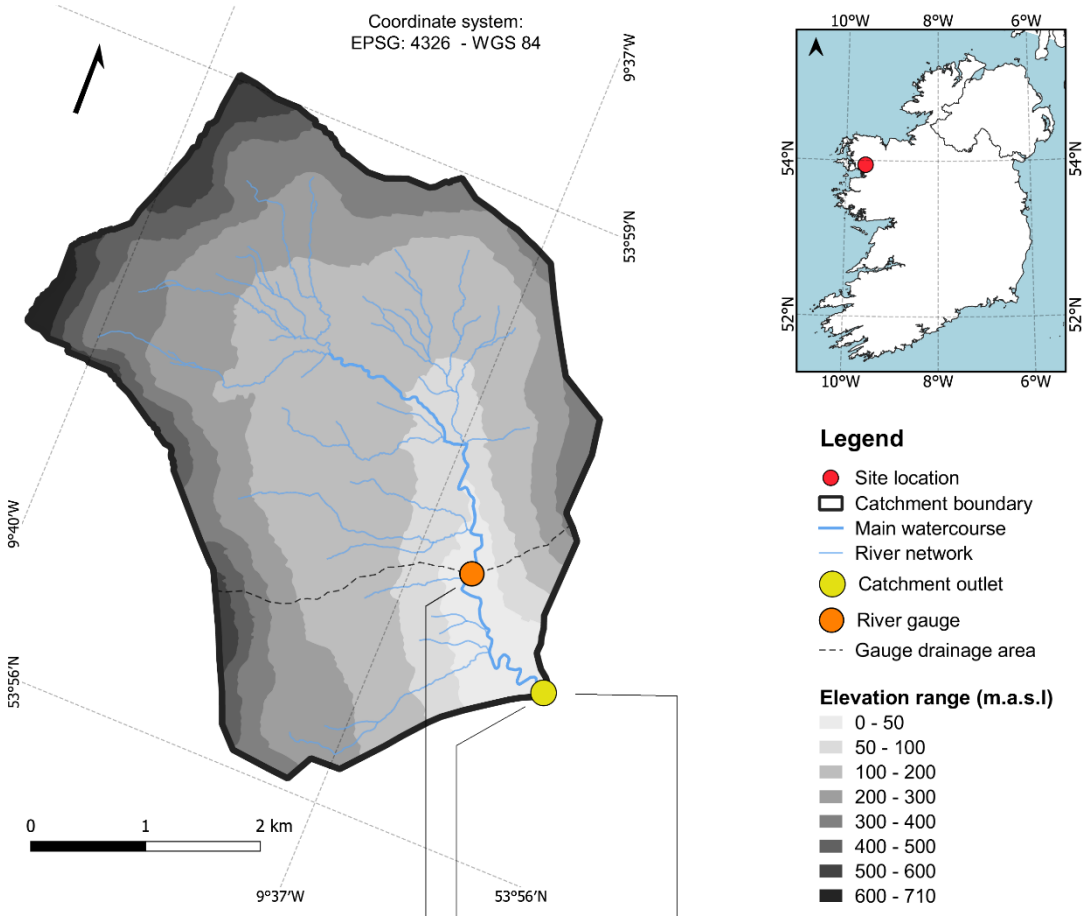


Figure 3. *The Glenamong River including its river network and topographical gradient. The catchment outlet and river gauge location are indicated in the figure.*

The dynamics of DOM in both study areas have been previously explored in Ryder et al. (2014) which identified soil temperature, river discharge and drought as important drivers in Feeagh catchment including the Glenamong River, and in Marcé et al. (2021) ,which showed human activities were significant drivers in Sau catchment. DOM in Feeagh is mostly driven by natural processes, while diffuse and point sources of organic matter and nutrients, e.g., wastewater effluents and agricultural runoff, are also important for Sau. Catchment hydrology is key for carbon transport into both study areas, and contributes

to a distinct seasonality related to climate (Ryder et al., 2014; Doyle et al., 2019; Marcé et al., 2021; Doyle et al., 2023). Feeagh has a wet temperate climate, with cooler air temperatures and higher rainfall levels that occur on more than 75% of days in the year. The variability of carbon inputs reflects the sensitivity of a peatland-dominated landscape, which exacerbates climate-induced carbon release from the catchment into the lake. In contrast, Sau has a Mediterranean climate, characterised by hot, dry summers and mild, wet winters, dictating water availability, thermal stratification, and organic matter fluxes in the reservoir.

2.2 Data availability on DOM

Data on DOM were available through two different proxies, and their used depended on the forecasting time scale and source water bodies. For the lakes, fluorescent dissolved organic matter (fDOM) was used as the target variable. For the river, dissolved organic carbon (DOC) was used as proxy of DOM. The availability and characteristics of these datasets are summarised below.

Daily surface fDOM values for the lake sites were derived from high-frequency measurements collected every two minutes. In Lough Feeagh, fDOM was measured at a depth of 0.9 m, while in Sau Reservoir an average value was calculated across the upper 0 to 5 m of the water column. All fDOM values were expressed in quinine sulphate units (QSU). The time series spanned from May 2012 to November 2019 for Feeagh and from February 2017 to March 2020 for Sau, with some gaps in both datasets. The availability of fDOM data determined the periods in which the modelling workflow could be trained and tested. In Feeagh, fDOM was measured using a UV fluorometer (Seapoint Sensors Inc., Exeter, NH, USA) and water temperature was obtained from a multi-parameter sonde. In Sau, fDOM and water temperature were measured using a digital smart sensor and a multi-parameter sonde. Raw fluorescence data at both sites were corrected for the effect of water temperature using sensor-specific correction relationships. Additional details on these corrections are provided in the supplementary material.

For the river site, DOC was estimated based on two different parameters that were available as complementary proxies: water colour and fDOM. Water colour-based DOC

data (2016 – 2023) were converted from weekly water colour measurements based on absorption and expressed as mg PtCo L^{-1}, via a linear regression relationship with DOC measured in the laboratory: DOC = 0.07 water colour + 2.7, r^2 = 0.88, n = 551 (Doyle et al., 2019). The water colour measurements were performed using a HACH Dr 2000 spectrophotometer at 455 nm (wavelength accuracy of ±2 nm) on water filtered through Whatman GF/C filters (1.2 µm pore size). CDOM-based DOC data (2020 – 2023) were converted from daily averaged high-frequency (raw data interval: 2-minute) CDOM data. These were first temperature-corrected and then converted to DOC using a linear regression relationship with DOC laboratory measurements: DOC = 0.13 CDOM – 6.46, r^2 = 0.48, n = 1570 (Jennings et al., 2020).The data were collected using a Seapoint UV fluorometer sensor (Seapoint Sensors Inc., Exeter, NH, USA) and river water temperature data was measured using a Hydrolab Quanta (OTT Hydrometry Ltd., Sheffield, UK). The Seapoint CDOM fluorometer used UV light emitting diodes (LEDs) as the CDOM excitation source (excitation wavelength of 370 nm CWL, 12 nm FWHM; emission wavelength of 440 nm CWL, 40 nm FWHM, where CWL is the centre wavelength and FWHM is the full width at half maximum wave height).

DBP data were not used in this deliverable because the modelling and forecasting workflows were specifically designed to predict DOM dynamics at the source, where DOM is generated and mobilised. In contrast, DBPs form mainly downstream at water treatment works (e.g., trihalomethanes, HAAs) or within distribution networks after chlorination. Since DBP formation depends strongly on treatment conditions, operational procedures and chlorine contact time, besides natural DOM dynamics in the raw source, these data fall outside the scope of a source-water forecasting workflow. However, the modelling outputs generated here can be used to infer DBP formation potential in settings where co-located DBP datasets exist, enabling future site-specific relationships between predicted DOM and DBP measurements (Pedregal-Montes et al., 2024).

Moreover, other initially proposed study sites that are smaller Irish group water schemes (GWP) have not been addressed in this deliverable. This was because in-situ monitoring is still ongoing, and the volume and duration of DOM data available so far are insufficient to reliably train or evaluate the workflows. Additionally, data acquisition procedures from

external companies managing the source-water and treatment-plant monitoring at these GWS sites are still being processed, aligned with national regulatory requirements.

2.3 Seasonal modelling and forecasting workflow

2.3.1 Simulating DOM in lakes

The seasonal workflow consisted of a five-step procedure applied independently to each lake site. First, all potential predictors of fDOM were compiled, including meteorological variables, soil conditions, hydrological outputs, lake model variables and a seasonal indicator. In total, 24 drivers were assembled for each site. Second, supervised machine learning models were trained using 85 percent of the available fDOM time series for each lake. The remaining 15 percent of the data were withheld as an independent testing period to evaluate predictive performance and to detect potential overfitting by comparing against the training results.

Third, the most influential drivers were identified using feature-importance metrics extracted from the machine learning models. Only drivers that exceeded an importance threshold of 5 percent were retained for further analysis. To examine how individual drivers influenced predictions, partial dependence plots and SHAP analyses were applied to this reduced set of variables.

Fourth, the workflow was used to generate two sets of simulations. The first used all drivers with importance values above the threshold, and the second used a subset of predictors limited to those available from global reanalysis datasets. This step allowed assessment of model performance under conditions that favour wider scalability. Predictive skill was evaluated using the coefficient of determination, Kling–Gupta efficiency and root mean square error. Finally, the workflow was applied to both lakes using the same modelling structure, enabling direct comparison between sites with contrasting catchments, hydrological regimes and DOM dynamics.

It is of note that the monitoring arrangement deployed in the Burrishoole catchment, which is important for both Lough Feeagh and the Glenamong River (study sites in this deliverable), was substantially strengthened through two earlier projects: the Water JPI PROGNOS project (2016–2019) and the JPI Climate WATExR project (2017–2020). These initiatives advanced integrated high-frequency monitoring of these water bodies

and the coupling of these observations with dynamic lake water-quality modelling frameworks, such as GLM (the lake model used in this work). In doing so, they enabled the development of continuous, quality-controlled datasets and modelling infrastructure that were essential for implementing the seasonal DOM workflow presented here, not only for the Irish site but also for the Spanish site.

Moreover, although the initial plan referenced the potential use of FLARE (Forecasting Lake and Reservoir Ecosystems) (Thomas et al., 2020; Carey et al., 2022), this system was not employed to generate the forecasts in our study lakes. FLARE is designed primarily for near-real-time, short-term forecasting, typically at daily to weekly lead times, and is tightly coupled to specific data-assimilation routines and driver requirements that differ from the objectives of this deliverable. In contrast, our seasonal framework required longer lead times (1–7 months) and the flexibility to investigate a broader set of environmental drivers, options that were better supported through the machine-learning approach adopted here. Nonetheless, FLARE played an important role in informing aspects of our modelling configuration, particularly in guiding the setup of GLM parameterisation, driver structuring and workflow organisation, drawing on practices developed under FLARE and similar modelling efforts.

2.3.2 Driver data

The input data for the workflow comprised 24 driving variables at each site. These were grouped into five categories: (1) meteorology, (2) soil, (3) process-based hydrological modelling, (4) process-based lake modelling and (5) Julian day. All input data variables, including their respective units and source, are displayed in Step 1 of Figure 4. Daily values of meteorology and soil variables were extracted from the European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis v5 (ERA5) dataset. This gridded dataset provides (pseudo) observations at a global scale with a spatial resolution of 0.25°. Eight meteorological variables (traditionally employed in water modelling studies) and soil temperature and soil moisture data at four depths (0-7 cm, 7-28 cm, 28-100 cm, and 100-255 cm) were extracted for the grid-cell which contained each water body.

Daily values of inflow discharge and inflow DOC concentration into each site were generated using the Generalised Watershed Loading Functions Model (GWLF) coupled with a DOC module (GWLF-DOC). The GWLF-DOC version and calibration strategy applied are described in Paíz et al. (2025). Calibration results can be found in the Supplementary Information. Daily values of five key lake variables (see Step 1, Figure 4) were obtained from the General Lake Model (GLM) (Hipsey et al., 2019) run for both sites. Calibration strategies applied are described in previous work (Mercado-Bettín et al., 2021; Paíz et al., 2025b). Calibration results can be found in the Supplementary Material. In addition, the cosine (to avoid an abrupt numerical change at every start of a year) of Julian day was included in the driver data inputs, given that seasonality is expected to influence DOM predictions.

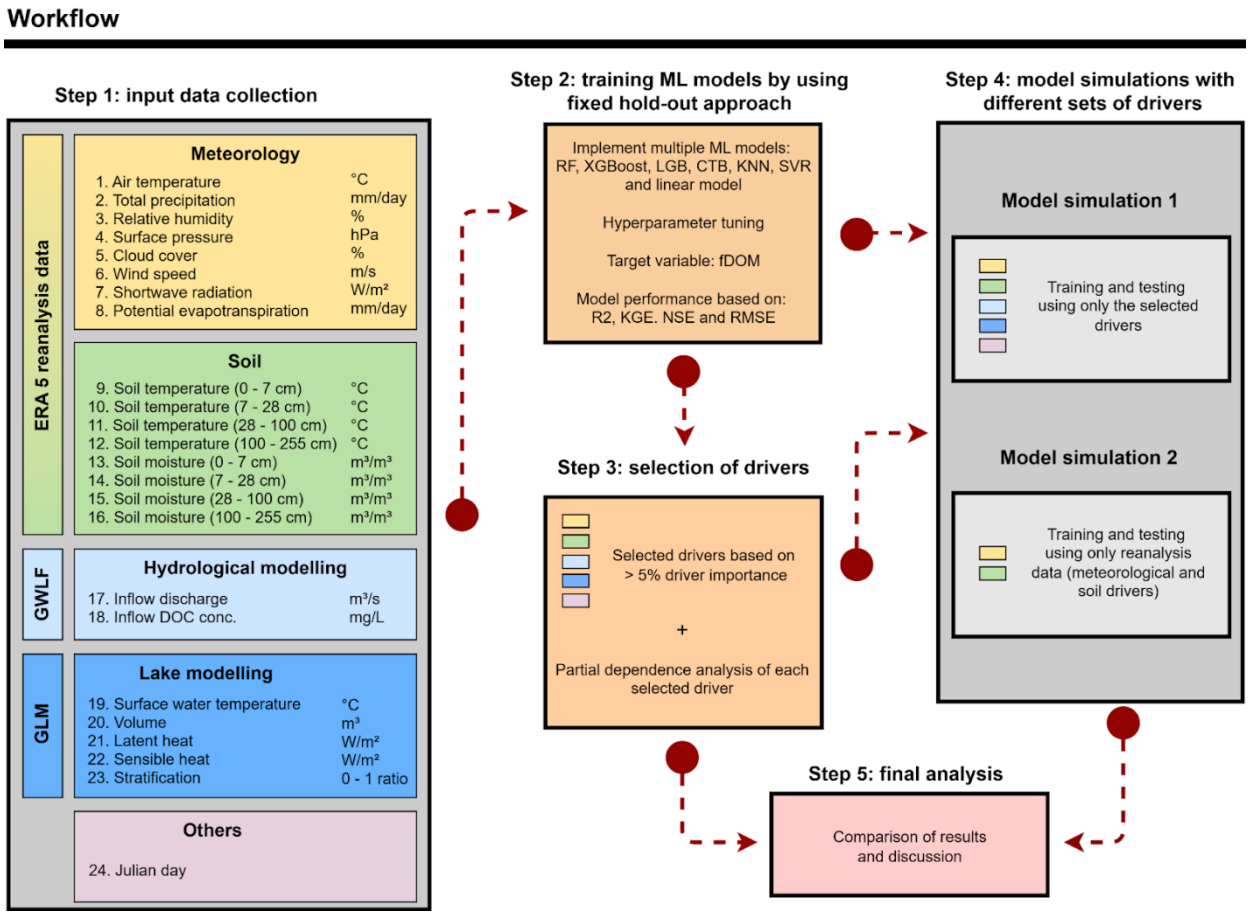


Figure 4. *Workflow implemented for simulating fDOM in both lake study sites.*

2.3.3 Supervised machine learning

Supervised ML models have advantages and limitations for time series prediction. In addition to capturing non-linear relationships typical in aquatic systems and water quality predictions, they provide flexibility to assess multiple drivers, temporal indicators and variables external to the system (Hollister et al., 2016; Regier et al., 2023). ML models do not require a fixed set of drivers to predict fDOM effectively, unlike process-based models, which typically rely on predefined inputs. Additionally, there is no need for parameter calibration but hyperparameter tuning, simplifying the modelling process. While some may argue that the lack of parameterization suggests a black box approach, supervised ML can provide insights into the potential drivers for predicting a target variable (Toming et al., 2020; Harkort and Duan, 2023).

However, due to the intrinsic autocorrelation in time series, for example when predicting DOM, these models tend to overfit when using out-of-bag samples during training. To overcome this issue, robust validation and training are required. Here, we used a hold-out period for validation during testing at both study sites. Prior to selecting this method, we compared it with two alternative validation methods using random forest: 5-fold cross-validation and rolling window cross-validation with a two-year training period, a one-year testing period and a window shift every 90 days. The comparison revealed that Feeagh exhibited more consistent model performance, with less overfitting between training and testing phases, across the different validation methods compared to Sau. This difference is likely attributable to the limited amount of available data at Sau.

Seven ML and statistical models were used to predict fDOM dynamics: Random Forest, eXtreme Gradient Boosting, Light Gradient Boosting, CatBoost, k-Nearest Neighbors, Support Vector Regression and a linear model. Random Forest, XGBoost, LGB and CatBoost can directly provide the importance of the features to predict the target variable; therefore, only these four models were used to select the most important drivers to predict fDOM. For this, the increase in node purity was used for Random Forest and the gain contribution of each feature for XGBoost, LGB and CatBoost. Further, hyperparameter tuning was implemented in all ML models to improve accuracy and generalisation. To extract the importance of the drivers, implement hyperparameter tuning and predict

fDOM, multiple R packages were used: caret, randomForest, xgboost, lightgbm, catboost, kknn and kernlab.

2.3.4 Partial dependence plots and SHAP analysis

To assess the influence of the most important drivers, we included partial dependence and SHapley Additive exPlanations (SHAP) plots, using the Random Forest and CatBoost models, respectively. The partial dependence plots illustrate the individual influence of each driver on fDOM predictions by varying the driver's values across its entire range while keeping all other drivers constant in their average value. The SHAP plots measure how much a single driver value contributes to moving the prediction away from the average value. The Y-axis has the input drivers ranked by overall importance from top to bottom, the X-axis has the SHAP value representing the impact on model output for a single prediction, each point is a single data instance and the color reflects the driver value (blue indicates low values, red indicates high values). To implement partial dependence plots and SHAP analysis, the pdp package in R and the shap package in Python were used.

2.3.5 Forecasting DOM in lakes

We developed a seasonal forecasting framework to predict monthly average concentrations of surface fDOM in lakes and reservoirs one to seven months ahead. The forecasting approach builds on the simulation workflow described above. For the forecasting, the same two sites were used, but the models were driven by a reduced set of predictors selected according to their relative importance and their availability in an open-access seasonal forecast product.

Seasonal forecasts were generated using only the Random Forest Regression model at both study sites to maintain consistency and comparability, and were driven by seasonal forecast product on single levels of ECMWF, provided through the Copernicus Climate Service. This product consists of a 1-degree global grid with sub-daily and daily meteorological and soil variables available from 1993 onwards (hindcasts from 1993 to 2016 and real-time forecasts from 2017 to present). Only those predictors available in this dataset (meteorological fields, soil temperature and soil moisture) were used to drive the seasonal forecasts.

Prior to forecasting, verification exercises were carried out to assess how well the seasonal drivers reproduced observed conditions. These comparisons were made against reanalysis-based observations of ERA5 from the simulation workflow and were also contextualised relative to climatology. Climatology was defined using two reference periods, 1990 to 2016 and 2017 to 2024, which allowed evaluation of how seasonal forecast products compared to average historical conditions. We used the Continuous Ranked Probability Skill Score (CRPSS) to assess forecast skill relative to a climatological baseline. CRPSS quantifies how much better (or worse) an ensemble forecast performs compared to climatology, with positive values indicating improved skill, values near zero indicating similar performance and negative values indicating performance worse than climatology. Following verification, seasonal forecasts of fDOM were generated and compared against observations, with forecast skill interpreted in line with these climatology-based assessments. This approach provided a structured way to understand how predicted conditions deviated from typical seasonal patterns and the degree to which the forecasts offered value beyond historical averages.

Finally, we examined how these forecasting outputs could support drinking-water utilities by identifying periods of elevated fDOM risk and by providing lead times that can help operational planning. The structure of this framework, based on open-access drivers and a reduced set of influential predictors, also offers the potential for scalability to other lake and reservoir systems where observational data may be limited.

2.4 Daily-scale modelling and forecasting workflow

2.4.1 Simulating DOM in rivers

Daily DOC dynamics in the Glenamong River were simulated using the Generalised Watershed Loading Functions Model with an integrated DOC module (GWLF-DOC). The model structure and parameterisation followed the approach previously established for this catchment (Paíz et al., 2025a). Daily observed meteorological inputs, including total precipitation and mean air temperature, were used as forcing data, and the model was run on a daily time step within VENSIM software. Climate observations from local meteorological stations (2013 – 2023), daily river discharge data (2013 to 2023) and two

datasets of estimated DOC concentration at the river outlet were obtained from the Marine Institute monitoring services (https://www.marine.ie).

The hydrological component of GWLF-DOC was calibrated and validated using measured river discharge data, with calibration performed for the period 2013 to 2018 and validation for 2019 to 2023. The DOC module was calibrated using the longer record of estimated DOC concentrations derived from water colour measurements (2016 to 2023). Calibration limits for the hydrological and DOC parameters followed previously defined parameter ranges based on earlier modelling studies (Paíz et al., 2025a). A two-year spin-up period was used to stabilise the DOC pools representing the five hydrologically connected carbon reservoirs: the deep saturated zone, the saturated zone, the unsaturated zone, DOC-producing surface areas excluding wetlands and wetlands. Model performance was evaluated based on root mean squared error (RMSE), the Nash-Sutcliffe coefficient (NSE) and the coefficient of determination (r^2) (Krause et al. 2005) between daily observed and simulated river discharge and DOC concentration.

2.4.2 Forecasting DOM in rivers

Short-term DOC forecasts were generated using a single ensemble-based forecasting workflow in which the long-term calibrated hydrological and DOC model parameters remained fixed. Prior to each forecast, only the five DOC initial conditions were updated through data assimilation. Forecasts were produced for every day during two full hydrological years (1 March 2021 to 1 March 2023), with each forecast consisting of 35 future daily DOC predictions. The workflow was forced using a 31-member ensemble from the Global Ensemble Forecasting System (0.5-degree NOAA GEFS), extracted for the grid cell containing the river catchment. This ensemble provided forecasts of total precipitation and mean air temperature out to 35 days ahead.

A five-month model spin-up (1 October 2020 to 28 February 2021) was run before producing forecasts. During this period, data assimilation was applied but no forecasts were generated. A seven-day assimilation window preceded each daily forecast (days −7 to 0), during which CDOM-based DOC estimates were assimilated. These DOC measurements, available in near real-time and independent of the long-term calibration

record, were used to update the DOC initial conditions. The assimilation was performed using a Kalman filter with a measurement noise standard deviation of 3.8 mg L^{-1}.

Following data assimilation, each GEFS ensemble member was used to independently force GWLF-DOC, resulting in an ensemble of 31 predictions for each forecast horizon from 1 to 35 days ahead. This ensemble provided a representation of uncertainty arising from both meteorological inputs and system dynamics.

2.4.3 Driver data

All inputs used in the forecasting workflow were assessed to characterise their variability and magnitude. These included the GEFS ensemble meteorological inputs, the modelled hydrological forcing used by the DOC module and the CDOM-derived DOC observations used for data assimilation. Variability was quantified using standard deviations, and magnitude was assessed using mean, median and 95th percentile values. Daily surface runoff and baseflow components were also estimated using a hydrograph separation method. Daily ensemble-mean meteorological forecasts were compared to climate station observations using root mean squared error, the Nash–Sutcliffe coefficient and the coefficient of determination. All forcing data were evaluated seasonally across the forecasting period using spring (March–May), summer (June–August), autumn (September–November) and winter (December–February) periods.

2.4.4 Forecast evaluation

Forecast performance was evaluated using root mean square error, the Nash–Sutcliffe efficiency coefficient and the coefficient of determination between measured DOC concentrations and the ensemble-mean predictions. Forecast uncertainty was calculated as the standard deviation of the 31 ensemble members. Evaluations were conducted seasonally and across forecast horizons of 1, 7, 14, 21, 28 and 35 days. Forecast reliability was assessed by determining the percentage of observations contained within selected forecast confidence intervals (10, 25, 50, 75, 90 and 95 percent). An ideally calibrated ensemble would contain a matching percentage of observations within each respective interval. Departures from these values indicated underconfidence or overconfidence in the forecast uncertainty (Thomas et al., 2020; Zwart et al., 2023).

To contextualise the added value of the forecasts, performance was compared to a baseline method commonly used for environmental variables: Site Persistence. Persistence forecasts assumed that DOC concentrations remained constant at the most recently observed value and were used to evaluate whether the process-based forecasting framework improved upon a simple and widely applied benchmark (Murphy, 1992; Lofton et al., 2023).

3. RESULTS

3.1 Seasonal-scale simulations, insights and predictions

3.1.1 Driver attribution

The most influential predictors of fDOM at each site were identified from all 24 potential drivers based on the 5% threshold of the variable importance extracted from the RF, XGB, LGB and CTB models. This resulted in eight influential drivers being identified for Feeagh and five for Sau (Figure 5), four of which were common to both sites.

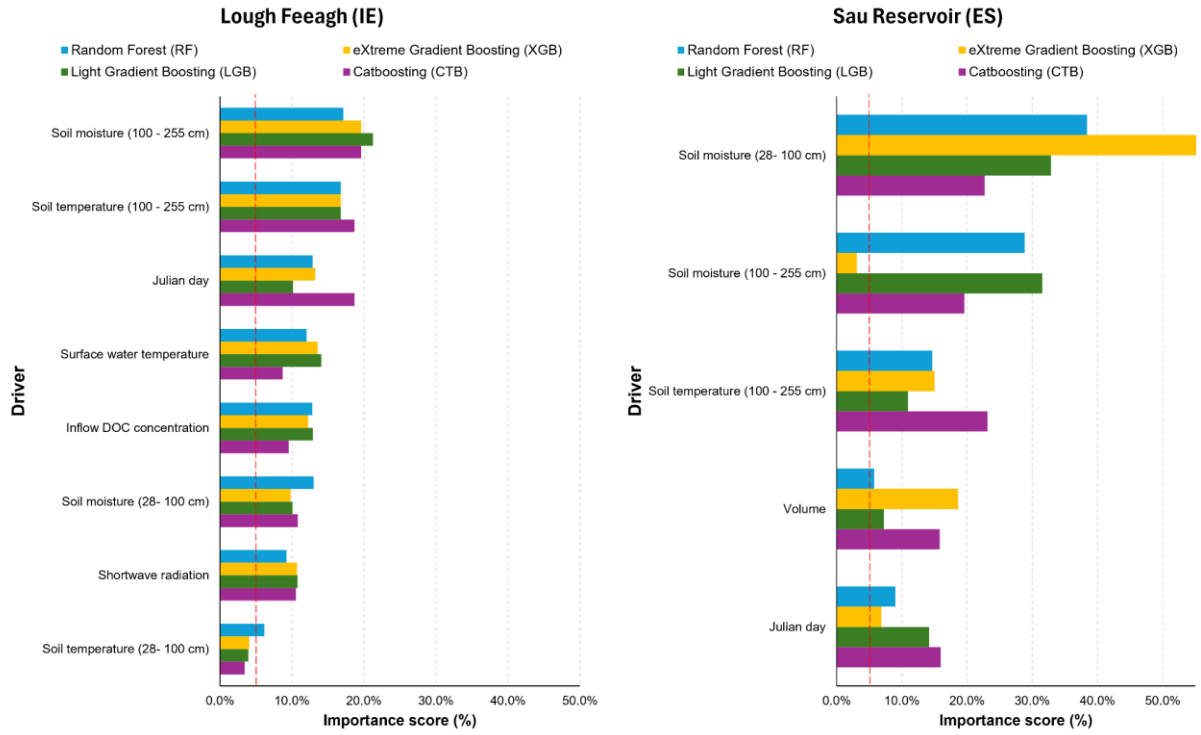


Figure 5. *Selecting influential drivers to predict fDOM. 24 drivers from various sources were used to train the four ML models that directly provide feature importance: Random Forest (RF), eXtreme Gradient Boosting (XGBoost), Light Gradient Boosting (LGB), CatBoost (CTB). These included 8 climate variables, 8 soil variables, 2 outputs from hydrologic and water quality modeling, 5 outputs from lake modeling, and the cosine of the Julian day. The most relevant drivers were identified based on their contribution to node purity or gain contribution, with only those exceeding a 5% variable importance being selected. For both case studies, the key drivers were soil temperature at 100–255 cm, soil moisture at 28–100 cm and 100–255 cm, and Julian day. Additionally, lake surface water temperature, inflow DOC concentration, and soil temperature at 28–100 cm were selected for Feeagh, while lake volume was selected for Sau.*

The variables for the deepest soil layer were relevant for both study sites. Soil temperature and soil moisture at 100-255 cm, were the most influential drivers for Feeagh. Similarly, for Sau, the deepest soil moisture driver was remarkably the most influential, while the deepest soil temperature was still important but less so than in Feeagh. Another key driver that was shared between Feeagh and Sau was Julian day.

Lake volume was only important for Sau, while solar radiation, the amount of carbon entering the water body (indicated by the DOC inflow concentration) and both soil moisture and temperature at 28-100 cm were only influential for Feeagh.

3.1.2 Partial dependency on selected drivers

Figure 6 introduces partial dependence plots and SHAP beeswarm plots for the most influential drivers selected in Figure 5, enabling the assessment of the individual effect of each driver on fDOM predictions. Seasonal patterns in fDOM concentrations were observed in both Feeagh and Sau, with higher values in winter and lower in summer, as reflected in the influence of Julian day. However, the key predictors and their effects differed substantially, shaped by contrasting catchment and climate characteristics. In Feeagh, where precipitation is relatively high and sustained year-round, resulting in a non-water-limited catchment, deep soil temperature (100–255 cm) was the dominant predictor, with fDOM increasing with temperature up to a threshold, beyond which a drop in the water table may counteract the effect. In addition, Feeagh showed minimal influence of mid-depth soil temperature (28-100 cm) and solar radiation on fDOM. This temperature-fDOM relationship was less relevant in Sau, where deep soil temperature (100-255 cm) had less explanatory power. Instead, fDOM dynamics in Sau, a water-limited catchment, were driven primarily by soil moisture at both 28-100 cm and 100-255 cm depths.

Water availability also shaped the role of other predictors differently across the two sites. For instance, surface water temperature in Feeagh showed a clear threshold behaviour, with fDOM increasing relatively linearly beyond 6.5ºC and stabilizing around 7.5ºC, while in Sau, water volume acted as a surrogate for fDOM production. Lower volumes corresponded to reduced fDOM values, which increase and stabilize beyond 146 hm^3. Interestingly, inflow DOC concentration has a slight influence on Feeagh but not on Sau, possibly due to hydro climatological processes controlling the relationships. These differences highlight how catchment water availability fundamentally alters the relative importance and behaviour of fDOM predictors.

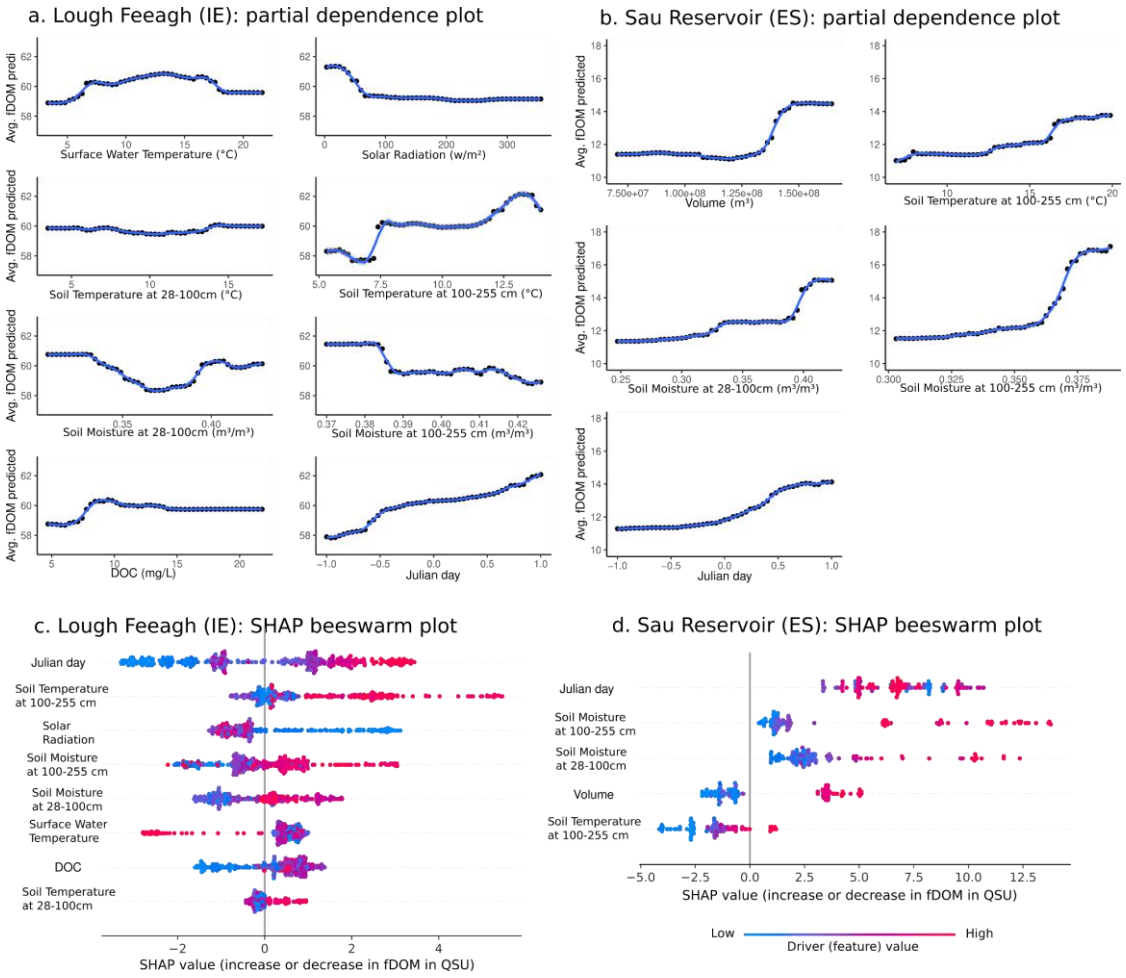


Figure 6. *Driver influence in predicting fDOM. It presents the partial dependence plots for (a.) Lough Feeagh and (b.) Sau Reservoir based on the random forest (RF) model, and SHAP beeswarm plots for (c.) Lough Feeagh and (d.) Sau Reservoir based on the CatBoost (CTB) model. In Feeagh, soil temperature at the deepest layer strongly influenced fDOM predictions, while in Sau, soil moisture at the deepest layer plays a significant role due to water availability constraints. In both cases, increases in these key drivers correspond to increases in fDOM. In both study sites, Julian day was a relevant factor, although the influence of seasonality on fDOM predictions was more evident in Feeagh.*

3.1.3 Simulating DOM using Supervised Machine Learning

Table 1 presents a comparative evaluation of the seven ML and statistical models employed to implement short-term forecasting of DOM. In Lough Feeagh, CatBoost (CTB) demonstrated the best overall performance with the highest R^2 (0.51) and KGE (0.69), and a relatively low RMSE (8.29). Similarly, in Sau Reservoir, CTB again has the highest R^2 (0.54) and KGE (0.66), and one of the lowest RMSE (7.11). While some models like RF and LM showed moderate performance at Feeagh, others such as SVR and XGB performed poorly in Sau, with SVR even having a negative KGE (-0.82), suggesting significant model bias. Overall, CatBoost (CTB) consistently outperformed all

other models across both sites, supporting its selection for fDOM prediction using the selected environmental drivers.

The results during the training phase (Supplementary Table S1) confirm that most ML models, especially XGB, RF, and KNN, had a high performance during training. However, the performance dropped in some models (e.g., XGB at Sau Reservoir) during the testing phases, underscoring the importance of evaluating models on independent test data to assess generalisability and overfitting. CatBoost (CTB), again, presented a more stable performance, showing slightly lower training metrics (especially in Feeagh) compared with the other ML models and better generalization when comparing with the metrics during testing, supporting its suitability for prediction.

Table 1. *Model comparison to predict fDOM using all selected drivers in both study sites. Statistic metrics (R^2, RMSE and KGE) were calculated to compare the performance of the models during the testing (hold-out) period. The table support the selection of the best model for each study site between Random Forest (RF), eXtreme Gradient Boosting (XGBoost), Light Gradient Boosting (LGB), CatBoost (CTB), k-Nearest Neighbors (KNN), Support Vector Regression (SVR) and linear model. Overall, the best performance was found in the CatBoost model. Supplementary Table S1 contains the results during the training period.*

	Lough Feeagh							**Sau Reservoir**						
Model/ Metric	**RF**	**XGB**	**LGB**	**CTB**	**LM**	**KNN**	**SVR**	**RF**	**XGB**	**LGB**	**CTB**	**LM**	**KNN**	**SVR**
R^2	0.37	0.41	0.47	0.51	0.4	0.36	0.34	0.53	0.11	0.45	0.54	0.45	0.33	0.63
RMSE	9.04	9.27	8.22	8.29	8.52	9.85	10.54	8.24	12.4	9.98	7.11	6.58	9.23	17.05
KGE	0.55	0.52	0.63	0.69	0.55	0.12	0.3	0.55	0.21	0.33	0.66	0.31	0.43	-0.82

3.1.4 Model prediction using all drivers compared to a reduced set.

Figures 7 and 8 present the fDOM prediction performance of the CatBoost model (best ML model overall) for Feeagh and Sau, respectively, using different input configurations. The models were trained on 85% of the time series (blue points) and tested on the remaining 15% using a hold-out approach (violet and green points). Two scenarios were compared: one using the most influential drivers (8 for Feeagh, 5 for Sau), and a second

using only a reduced subset of reanalysis-based and easily accessible drivers, specifically soil temperature, soil moisture, and Julian day.

For both lakes, the reduced driver models showed only a modest decline in predictive performance during the testing period. For example, in Feeagh, the model using all influential drivers achieved R^2 = 0.51 and KGE = 0.69, while the reduced model still attained R^2 = 0.48 and KGE = 0.67. Similarly, in Sau, the full model scored R^2 = 0.54 and KGE = 0.66, whereas the reduced model maintained a comparable R^2 = 0.50 and KGE = 0.65. Although the training performance was higher in Sau compared with the testing performance, indicating potential overfitting, the CatBoost model provided informative and generalizable predictions in both study sites.

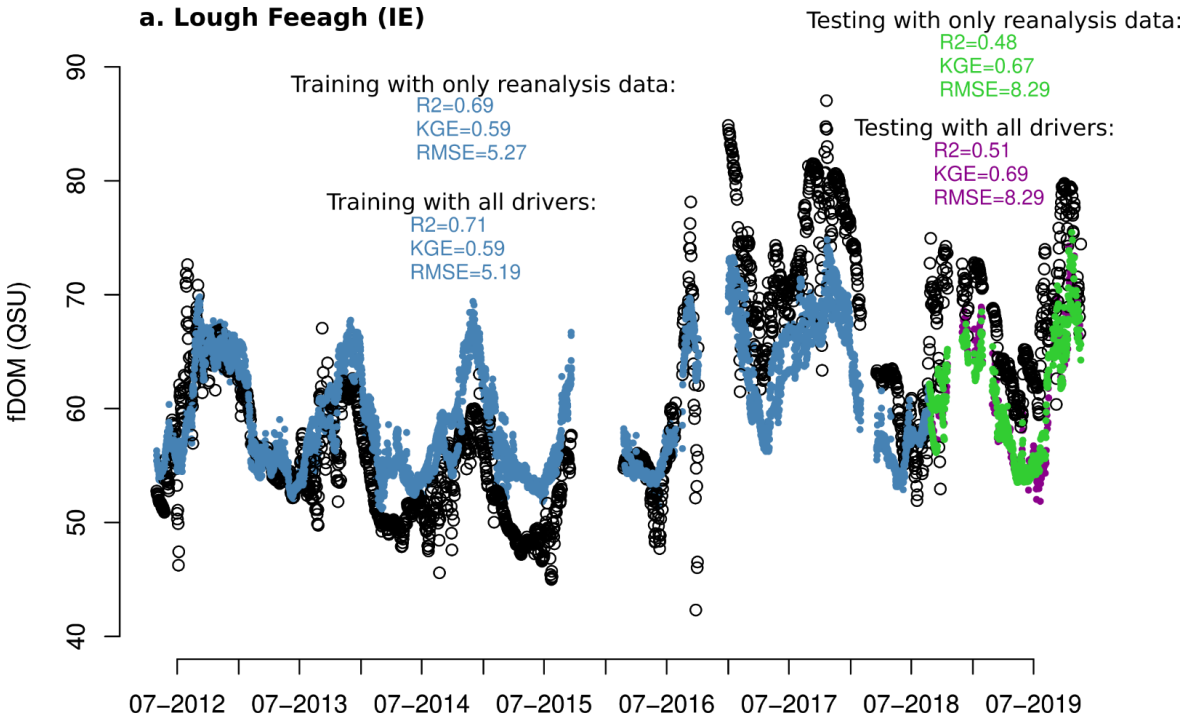


Figure 7. *fDOM predictions obtained from the CatBoost model using different driver sets. Here the training (blue, 85% of the time series) and testing (violet or green, 15% of the time series) periods are shown. The training and testing periods are completely independent and follow the hold-out period method, and model performance metrics (R^2, KGE, and RMSE) are calculated for both periods. a. Training and testing results for Feeagh: using the 8 most influential drivers (violet) and using only easily accessible and reanalysis data (soil temperature, soil moisture and Julian day) from those influential drivers (green).*

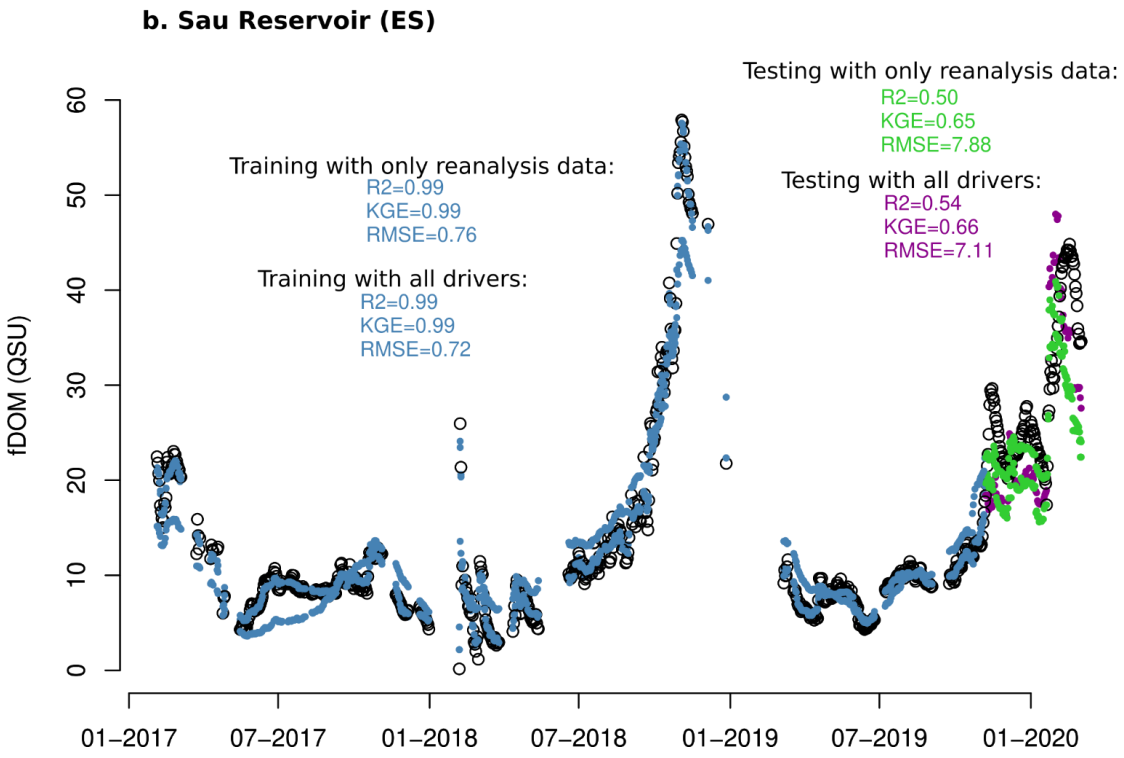


Figure 8. *fDOM predictions obtained from the CatBoost model using different driver sets. Here the training (blue, 85% of the time series) and testing (violet or green, 15% of the time series) periods are shown. The training and testing periods are completely independent and follow the hold-out period method, and model performance metrics (R^2, KGE, and RMSE) are calculated for both periods. b. Training and testing results for Sau: using the 5 most influential drivers (violet) and using only easily accessible and reanalysis data (soil temperature, soil moisture and Julian day) from those influential drivers (green).*

3.1.5 Seasonal forecasts

Verification of the seasonal forecast product against observations was carried out for all meteorological and soil variables used to drive the fDOM forecasts. The variables were: air temperature, total precipitation, soil moisture level 1, soil moisture level 2, soil moisture level 3, soil moisture level 4 and soil temperature level 1 (see Figure 4). The seasonal forecast fields showed a relative high correlation with reanalysis-based observations (ERA-5), indicating that the large-scale meteorological signals required for prediction were adequately captured. Across variables, the Continuous Ranked Probability Skill Score (CRPSS) was consistently greater than zero with few exceptions, demonstrating added value compared to using statistical average climatological conditions alone. These positive CRPSS values indicated that the seasonal forecast product had skill relative to climatology and therefore was suitable to drive seasonal-scale fDOM predictions.

Examples of the verification results for Sau Reservoir are shown in Figures 9 to 12. Verification figures for the remaining variables and for both study sites are provided in the Supplementary Material.

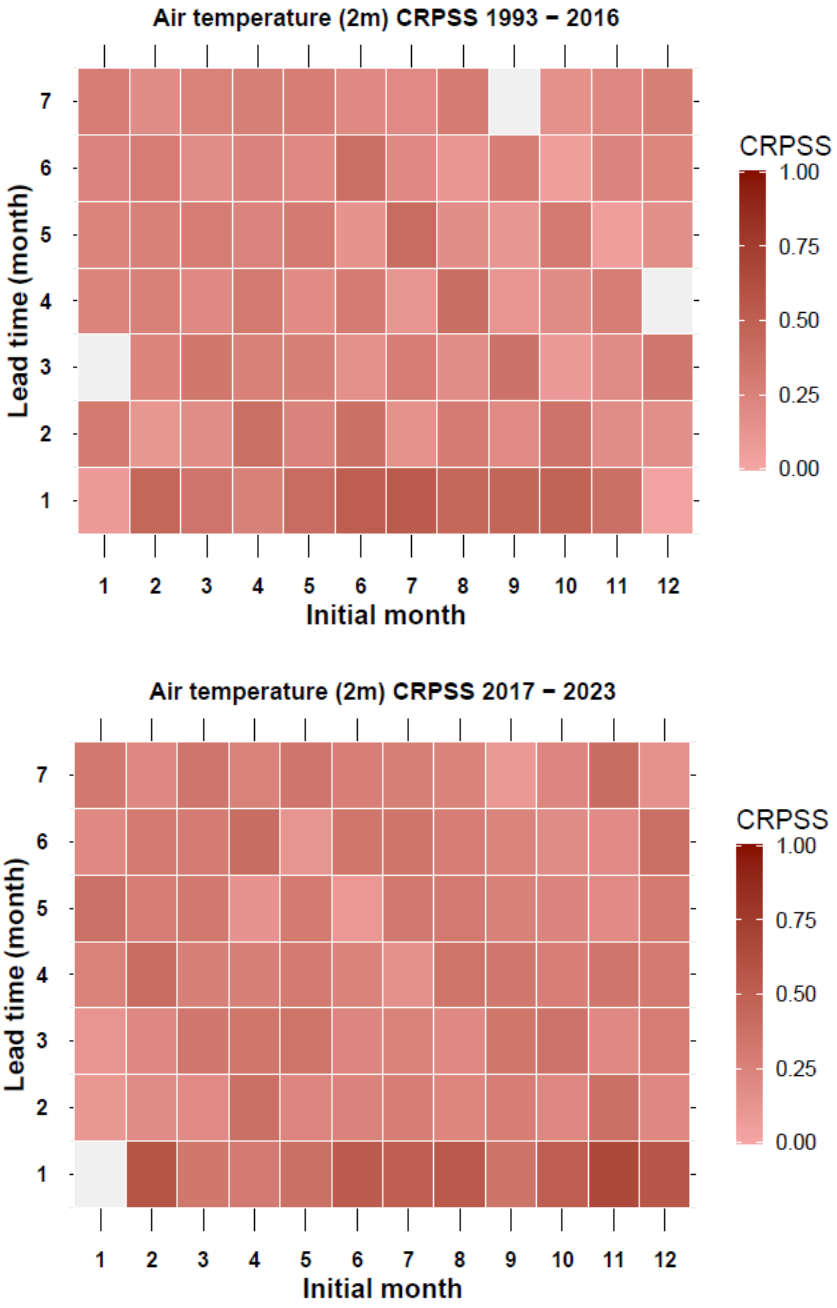


Figure 9. *Verification of the seasonal forecast product for 2-m air temperature using CRPSS for the periods 1993–2016 (upper panel) and 2017–2023 (lower panel) in Sau Reservoir catchment. CRPSS values above zero indicate added skill relative to climatology. Values are shown as a function of initial forecast month (x-axis) and forecast lead time in months (y-axis).*

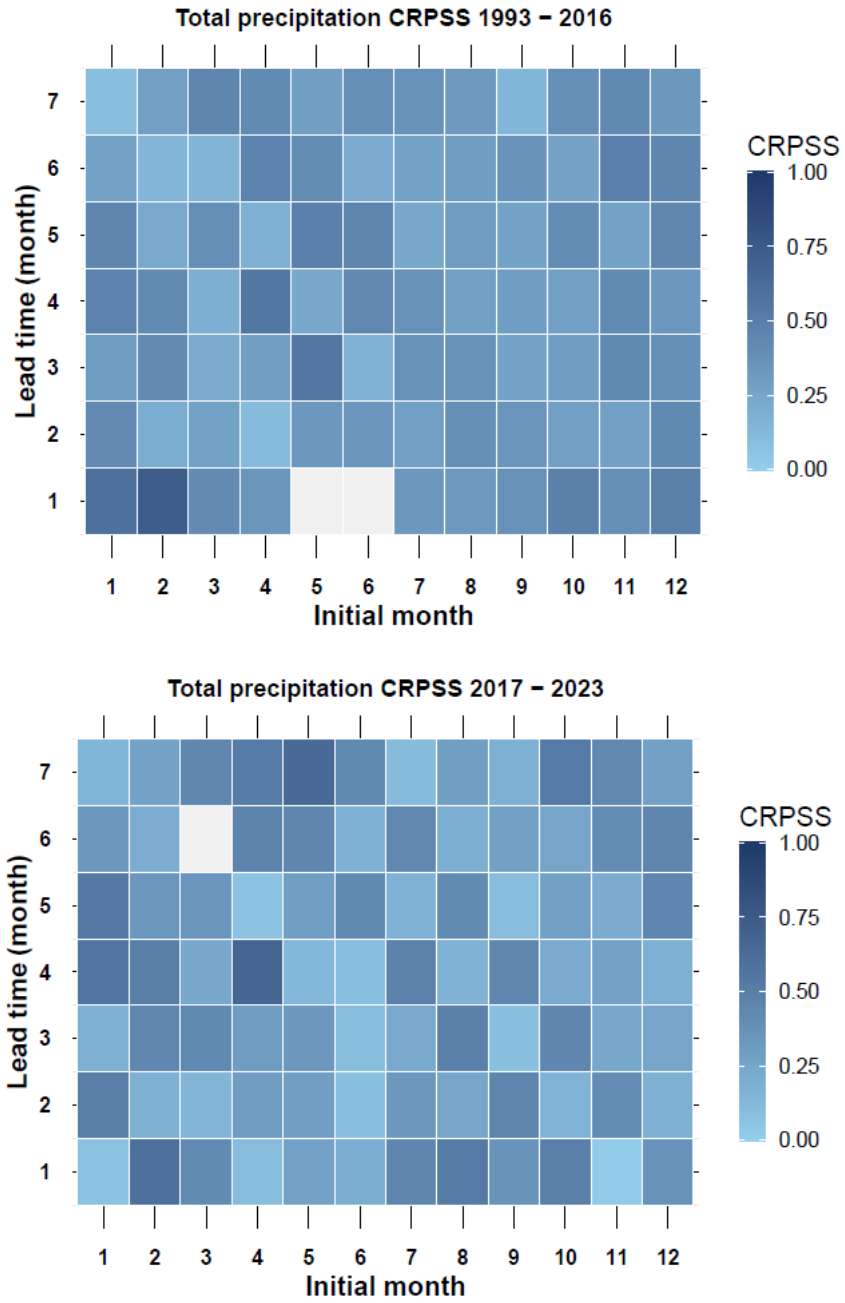


Figure 10. *Verification of the seasonal forecast product for total precipitation using CRPSS for 1993–2016 (upper panel) and 2017–2023 (lower panel) in Sau Reservoir catchment. CRPSS values above zero indicate that the seasonal forecast product outperforms climatological averages. Results are displayed across initial forecast months (x-axis) and forecast lead times from one to seven months (y-axis).*

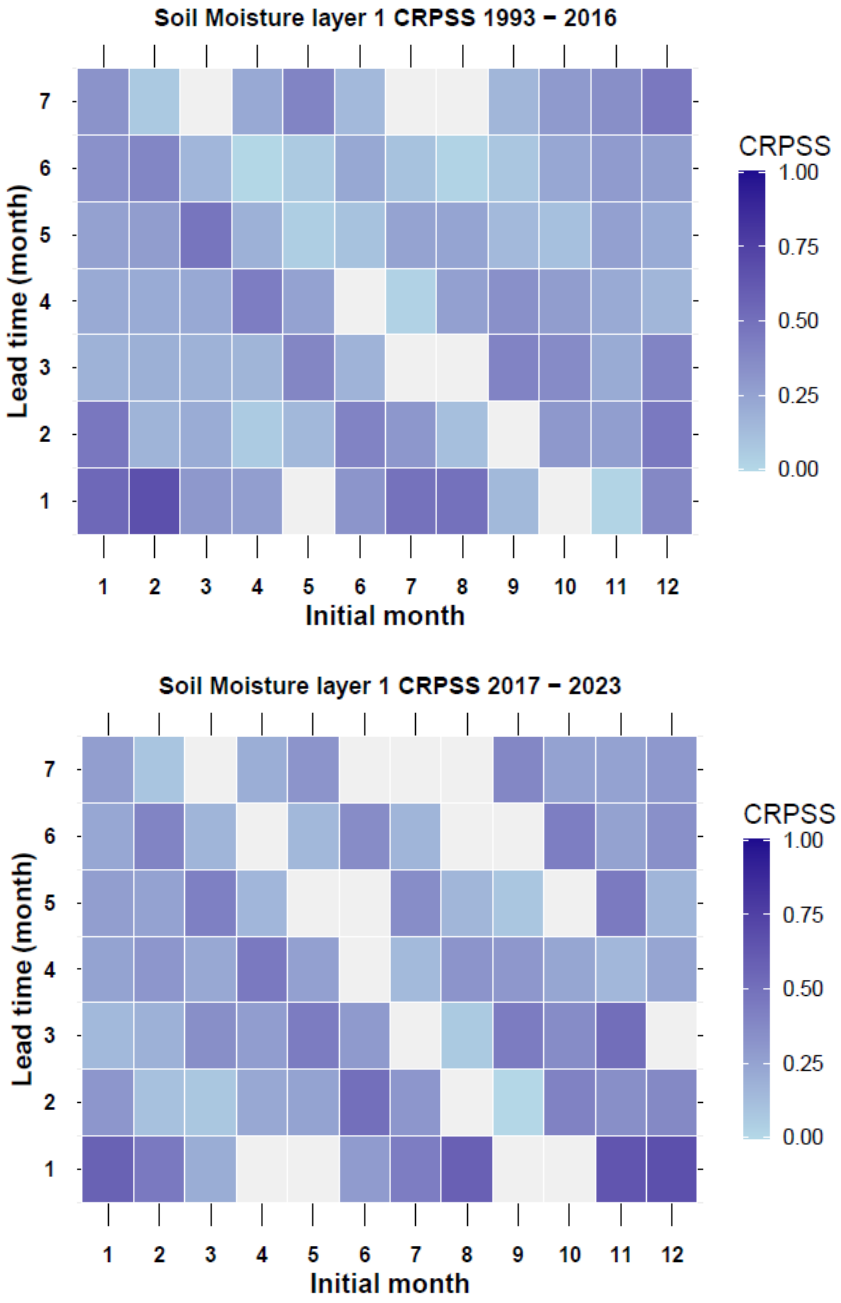


Figure 11. *Verification of the seasonal forecast product for soil moisture (Layer 1) using CRPSS for 1993–2016 (upper panel) and 2017–2023 (lower panel) in Sau Reservoir catchment. Positive CRPSS values indicate improved performance relative to climatology. The heatmaps show forecast skill across initial forecast months (x-axis) and forecast lead times (y-axis).*

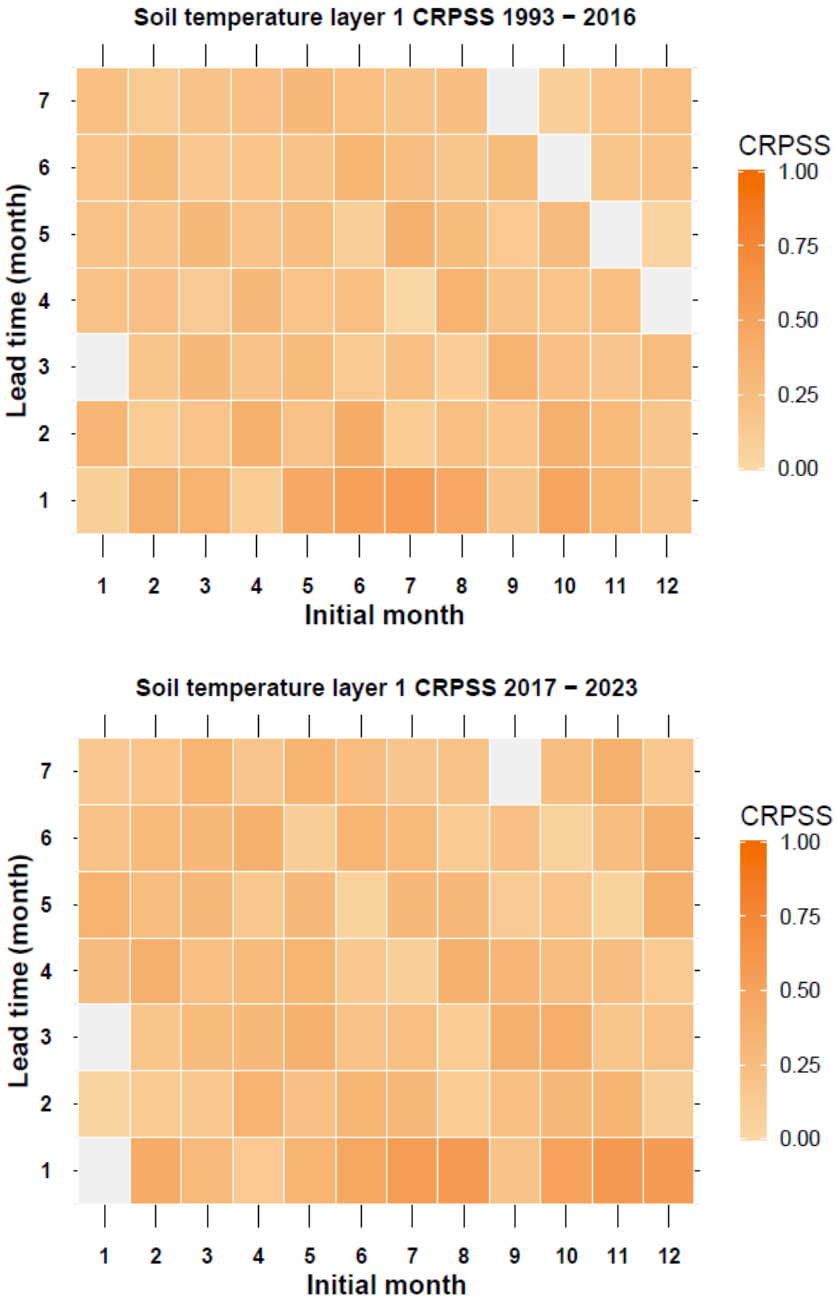


Figure 12. *Verification of the seasonal forecast product for soil temperature (Layer 1) using CRPSS for 1993–2016 (upper panel) and 2017–2023 (lower panel) in Sau Reservoir catchment. CRPSS values higher than zero indicate added value beyond climatology. Skill scores are presented for all initial forecast months (x-axis) and lead times up to seven months (y-axis).*

Seasonal fDOM forecasts were generated using the reduced predictor set identified in the simulation workflow and were driven by the bias-adjusted seasonal forecast variables. The predictor fields were first corrected using a quantile-based statistical adjustment derived from ERA-5 to ensure consistency between the long-term historical distributions and the seasonal forecast inputs. Following this correction, fDOM forecasts were produced for monthly averages one to seven months ahead.

Forecast results for the two study sites are shown in Figure 13 for Lough Feeagh and Figure 14 for Sau Reservoir. In both lakes, the forecasts captured the broad seasonal evolution of fDOM and demonstrated skill relative to climatology, with few exceptions which were generally at longer horizons. It is of note that some skill is just slightly better than climatology, although this was consistent in both sites. Across lead times, the majority of predictions remained within range of seasonal variability and provided meaningful anticipatory information (indicated by the CRPSS performance) beyond what could be inferred from climatological averages.

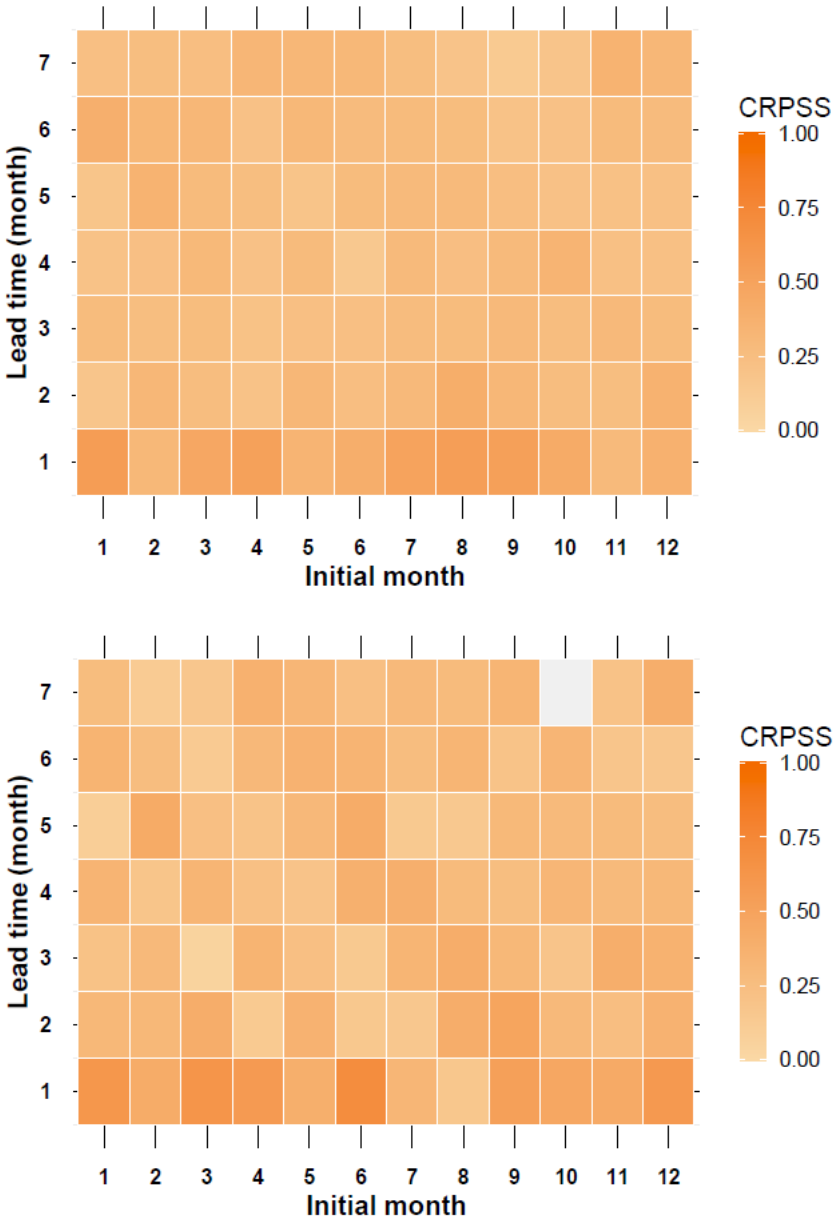


Figure 13. *Verification of the seasonal forecasts of fDOM monthly average concentration predictions using CRPSS for 1993–2016 (upper panel) and 2017–2023 (lower panel) in Lough Feeagh. CRPSS values higher than zero indicate added value beyond climatology. Skill scores are presented for all initial forecast months (x-axis) and lead times up to seven months (y-axis).*

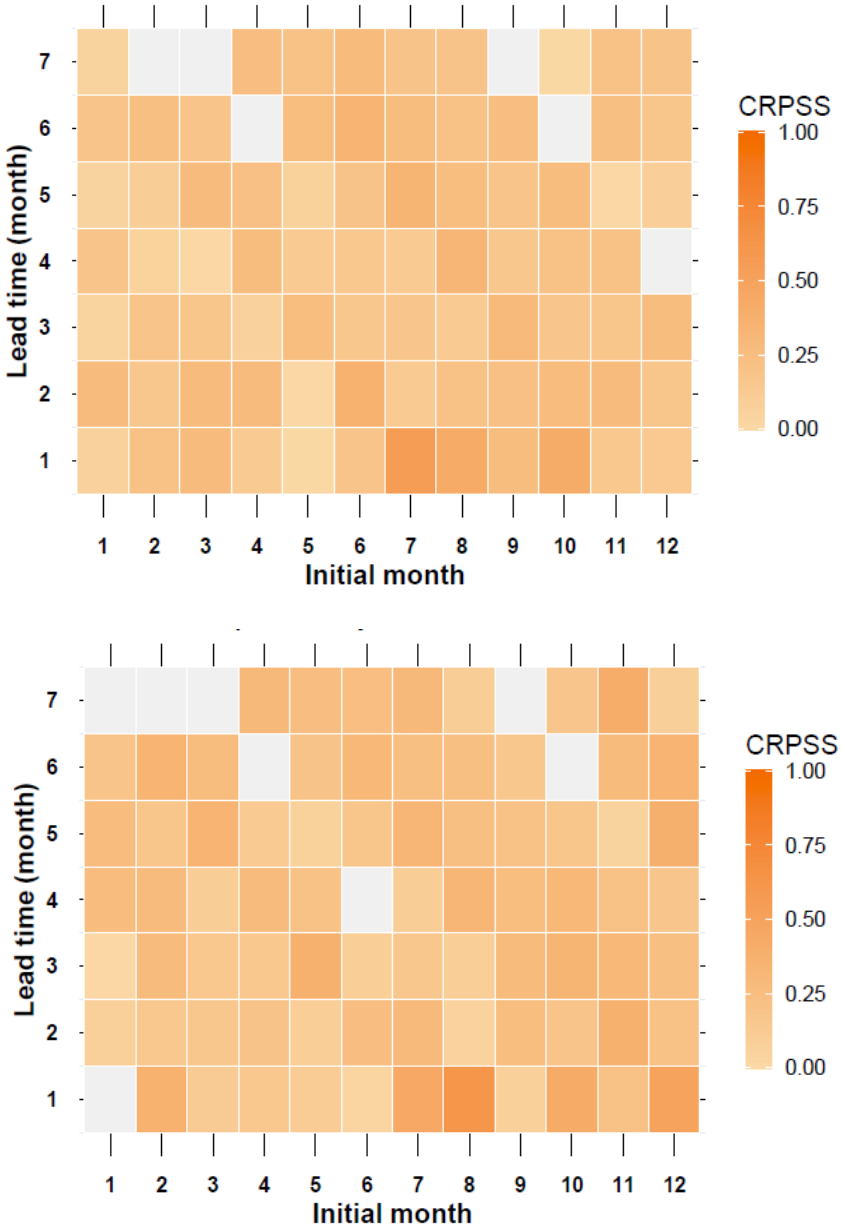


Figure 14. *Verification of the seasonal forecasts of fDOM monthly average concentration predictions using CRPSS for 1993–2016 (upper panel) and 2017–2023 (lower panel) in Sau Reservoir. CRPSS values higher than zero indicate added value beyond climatology. Skill scores are presented for all initial forecast months (x-axis) and lead times up to seven months (y-axis).*

3.2 Daily-scale simulations, insights and predictions

3.2.1 Process-based model performance

The optimisation of GWLF-DOC produced a good daily fit for both hydrology and DOC. For discharge, RMSE, NSE and r^2 were 7.44 mm day^{-1}, 0.64 and 0.64 during calibration, and 7.36 mm day^{-1}, 0.60 and 0.60 during validation (Figure 15). For DOC concentration, RMSE, NSE and r^2 during calibration were 2.49 mg L^{-1}, 0.45 and 0.47, respectively. Calibrated hydrological and DOC parameter sets and their ranges are listed in Table 2.

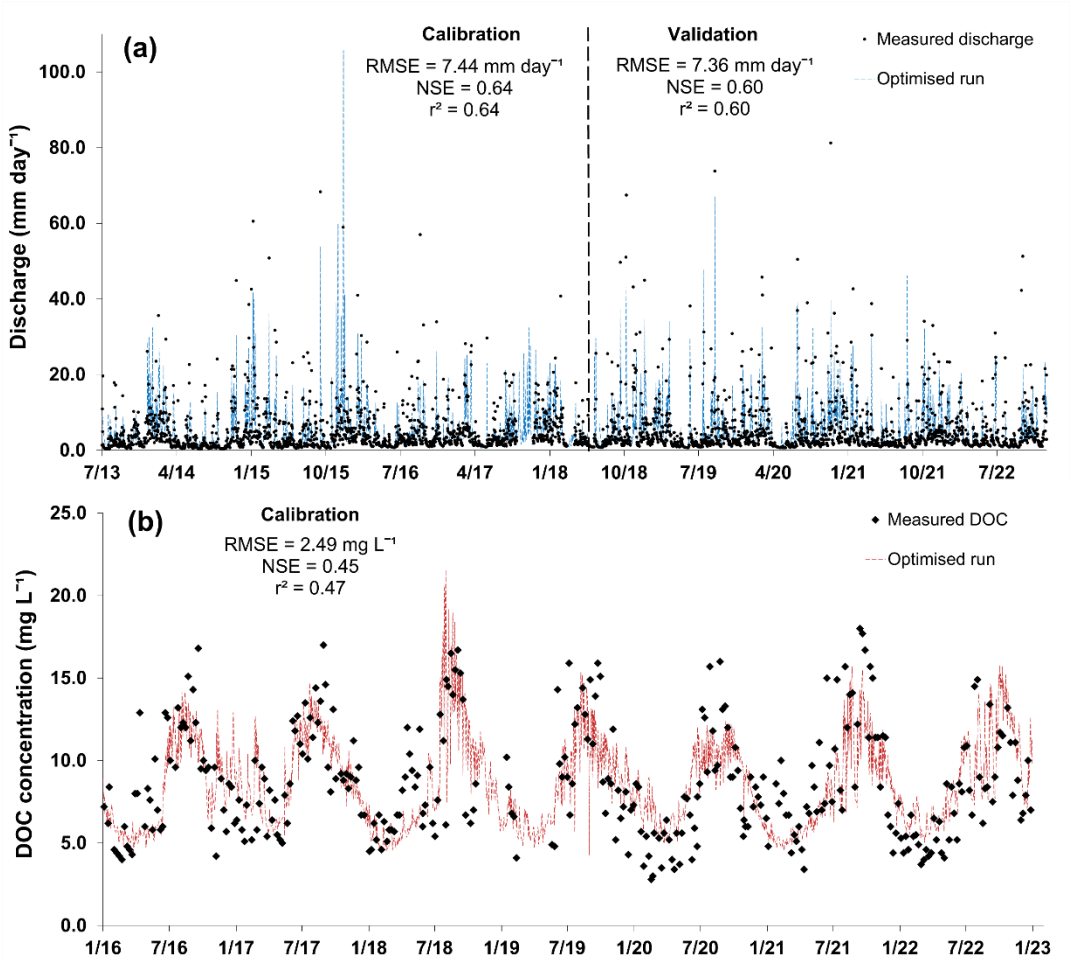


Figure 15. *Observed and simulated daily river discharge during the model calibration and validation of the hydrology module of GWLF-DOC (a), and observed and simulated daily DOC concentration during the model calibration of the DOC module of GWLF-DOC (b) for the Glenamong catchment.*

Table 2. *Observed and simulated daily river discharge during the model calibration and validation of the hydrology module of GWLF-DOC (a), and observed and simulated daily DOC concentration during the model calibration of the DOC module of GWLF-DOC (b) for the Glenamong catchment.*

Model calibration results			
Hydrology Module			
Model parameter	**Units**	**Calibration range**	**Calibrated value**
Precipitation correction factor	*factor / day*	*0.50 - 1.50*	1.08
River channel flow coeff.	*coeff. / day*	*0.01 - 1.00*	0.91
Melt coeff.	*cm / °C / day*	*0.05 - 0.60*	0.05
Curve adj. number	-	*-10.0 to 10.0*	5.50
Soil water capacity	*cm*	*12.00 - 20.00*	13.27
Baseflow capacity	*cm*	*2.50 - 8.00*	5.11
Slow recess coeff.	*coeff. / day*	*0.001 - 0.100*	0.010
Recess coeff.	*coeff. / day*	*0.005 - 0.030*	0.159
DOC Module			
Model parameter (DOC production)			
Activation energy (Ea)	*kJ / gC*	*3.0 - 7.0*	3.0
Aerobic decomp. rate (b)	-	*1.0 - 7.5*	1.0
Anaerobic decomp. rate (a)	*gC / m² / day*	*0.005 - 0.15*	0.096
Anaerobic decomp. rate wetlands (aw)	*gC / m² / day*	*0.005 - 0.15*	0.014
Model parameter (DOC washout - runoff)			
Surface carbon loss coeff. (Rcl)	*coeff. / day*	*0.01 - 1.00*	0.17
Surface flow cap (Sf cap)	*cm*	*0.01 - 6.00*	2.47
Model parameter (DOC washout - baseflow)			
Baseflow ratio (Br)	*fast- to slow-groundwater ratio / day*	*0.01 - 1.00*	0.87
Fast groundwater carbon loss coeff. (fgw)	*coeff. / day*	*0.001 - 1.000*	0.168
Slow groundwater carbon loss coeff. (sgw)	*coeff. / day*	*0.001 - 1.000*	0.002
Initial conditions - carbon pools			
Deep saturated zone	gC / m²	*0.0 - 28.5*	25.6
Saturated zone	gC / m²	*0.0 - 1.2*	0.4
Unsaturated zone	gC / m²	*0.0 - 1.2*	1.2
DOC-producing soils that are not wetlands	gC / m²	*0.0 - 2.9*	0.0
Wetlands	gC / m²	*0.0 - 1.5*	0.0

3.2.2 Forcing data assessment

Seasonal patterns in the weather forecasts, modelled hydrology and observed DOC used for data assimilation were consistent with a temperate site (Figure 16). Forecast mean air temperatures peaked in summer and were lowest in winter, with the largest variability in spring and autumn. Daily precipitation totals were highest in autumn, followed by winter, and lowest in summer. Against local observations, temperature forecasts showed RMSE 1.48–2.49 °C, NSE 0.87–0.53 and r^2 0.95–0.65 across the horizon. Precipitation skill was lower, with RMSE 6.09–8.10 mm day^{-1}, NSE 0.21 to −0.14 and r^2 0.22–0.00. Modelled daily catchment discharge used to drive the DOC module was largest in autumn (mean 2.9 mm day^{-1}; 95th percentile 7.9 mm day^{-1}) and winter (mean 2.5 mm day^{-1}; 95th percentile 7.8 mm day^{-1}), and lower in spring and summer (means 1.8 mm day^{-1}). Variability was greatest in autumn and winter. Mean baseflow contributions ranged from 65 percent in spring to 47 percent in winter, complementing surface runoff contributions from 35 to 53 percent (Table 3).

Observed daily DOC used for data assimilation was highest in summer (mean 10.9 mg L^{-1}; 95th percentile 15.2 mg L^{-1}) and lowest in winter (mean 5.4 mg L^{-1}; 95th percentile 9.0 mg L^{-1}). Variability was greatest in spring (standard deviation 3.6) and lowest in winter (standard deviation 2.1) (Figure 16d; Table 3).

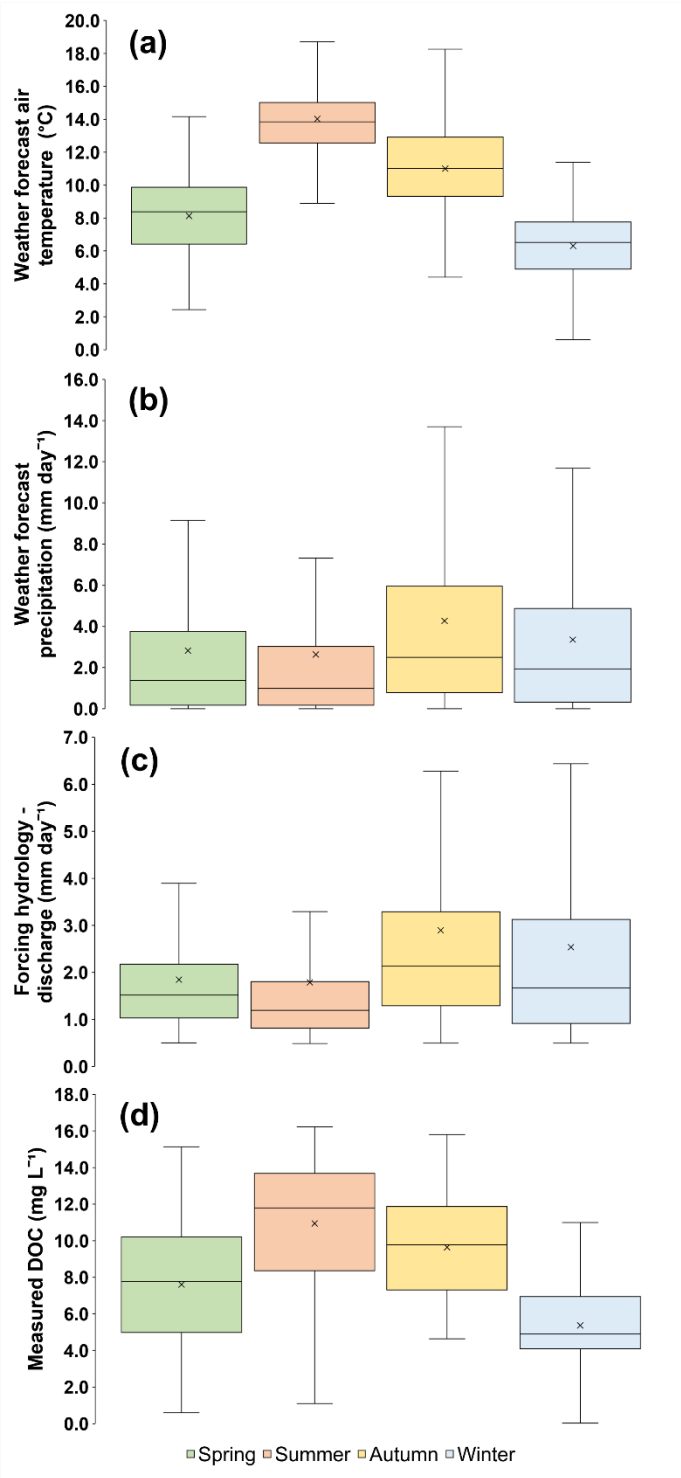


Figure 16. *Boxplot distributions of the weather forecast product input data for GWLF-DOC (i.e., daily forecasted catchment precipitation and mean air temperature) (a and b), modelled daily catchment discharge (c) and DOC observations used for data assimilation (d) over the meteorological seasons in the forecasting period (1st March 2021 – 1st March 2023).*

Table 3. *Magnitudes and variability across meteorological seasons of all forcing data used to generate DOC concentration forecasts i.e., the weather forecast input data (i.e., daily forecasted total precipitation and mean air temperature), the modelled forcing hydrology (daily catchment discharge, surface runoff and baseflow), and the DOC observations used for data assimilation during the entire forecasting period.*

Season	Spring	Summer	Autumn	Winter
Weather forecast product - air temperature				
Daily air temperature range (°C)	2.4 - 14.2	8.6 - 21.6	4.4 - 18.5	0.1 - 11.4
Mean daily air temperature (°C)	8.1	14.0	11.0	6.3
Median daily air temperature (°C)	8.4	13.8	11.0	6.5
95th percentile daily air temperature (°C)	11.8	18.0	14.9	9.6
Standard deviation (°C day^{-1})	2.4	2.2	2.6	2.2
Weather forecast product - precipitation				
Daily precipitation range (mm day^{-1})	0.0 - 39.7	0.0 - 44.7	0.0 - 52.4	0.0 - 37.5
Mean daily precipitation (mm day^{-1})	2.8	2.6	4.3	3.4
Median daily precipitation (mm day^{-1})	1.4	1.0	2.5	1.9
95th percentile daily precipitation (mm day^{-1})	10.8	11.7	14.1	11.7
Standard deviation (mm day^{-1})	3.9	4.3	5.2	4.2
Mean monthly precipitation (mm $month^{-1}$)	86.1	80.6	128.1	100.5
Mean precipitation (mm $season^{-1}$)	258.3	241.7	384.2	301.4
Forcing hydrology - river discharge				
Mean daily discharge (mm day^{-1})	1.8	1.8	2.9	2.5
Median daily discharge (mm day^{-1})	1.5	1.2	2.1	1.7
95th percentile daily discharge (mm day^{-1})	4.4	5.5	7.9	7.8
Standard deviation (mm day^{-1})	1.4	2.1	2.9	2.6
Forcing hydrology - surface runoff				
Mean daily surface runoff (mm day^{-1})	0.8	0.7	1.5	1.2
Mean daily contribution to river discharge (%)	35%	46%	50%	53%
Forcing hydrology - baseflow				
Mean daily baseflow (mm day^{-1})	0.2	0.2	0.7	0.5
Mean daily contribution to river discharge (%)	65%	54%	50%	47%
Measured DOC				
DOC concentration range (mg L^{-1})	0.6 - 15.1	1.1 - 16.2	4.6 - 15.8	0.1 -11.0
Mean DOC concentration (mg L^{-1})	7.6	10.9	9.6	5.4
Median DOC concentration (mg L^{-1})	7.8	11.8	9.8	4.9
95th DOC concentration (mg L^{-1})	12.9	15.2	14.1	9.0
Standard deviation (mg L^{-1} day^{-1})	3.6	3.4	2.8	2.1

3.2.3 Daily-scale forecasts

Forecasts were generated daily with 35-day horizons and evaluated against observations (Figures 17 to 19). Divergence between forecasts and observations increased with lead time, as expected, and newly available observations were assimilated as the forecast window advanced. During the 7-day data assimilation window preceding each forecast, only the five initial conditions representing catchment carbon pools were adjusted, while hydrological and DOC model parameters remained fixed at their long-term calibrated values. Initial conditions exhibited meaningful variability across forecasts, with larger adjustments typically required for the saturated zone and surface-area pools during wetter seasons, and for the unsaturated and deep saturated zones during drier conditions. This behaviour is consistent with seasonal shifts in hydrologic connectivity and washout potential.

Performance depended on season and lead time. In spring, the workflow achieved RMSE 1.77–4.56 mg L^{-1}, NSE 0.75 to −0.56 and r^2 0.77–0.07 across 1–35 days. Skill was strongest at short leads, particularly the first 3 days, where RMSE was below 2.26 mg L^{-1}, NSE above 0.59 and r^2 above 0.64. In summer, performance was similar, with RMSE 2.59–4.33 mg L^{-1}, NSE 0.45 to −0.55 and r^2 0.62–0.13, again with best metrics at short leads, for example 3-day horizons with RMSE below 3.13 mg L^{-1}, NSE above 0.19 and r^2 above 0.50. In autumn and winter, skill decreased relative to spring and summer, but the workflow retained useful short-lead performance. In autumn, the first 4 days achieved RMSE below 2.61 mg L^{-1}, NSE above 0.15 and r^2 above 0.35. In winter, the first 3 days achieved RMSE below 1.93 mg L^{-1}, NSE above 0.21 and r^2 above 0.35 (Figures 18 and 19).

Reliability analysis using central confidence intervals indicated an overconfident ensemble, with observed frequencies within nominal intervals lower than ideal across horizons and seasons. Forecast uncertainty increased gradually with lead time in all seasons (Figure 19). Standard deviation of the forecast ensemble ranged from 0.4 mg L^{-1} at 1 day ahead in winter to 1.6 mg L^{-1} at 35 days ahead in autumn.

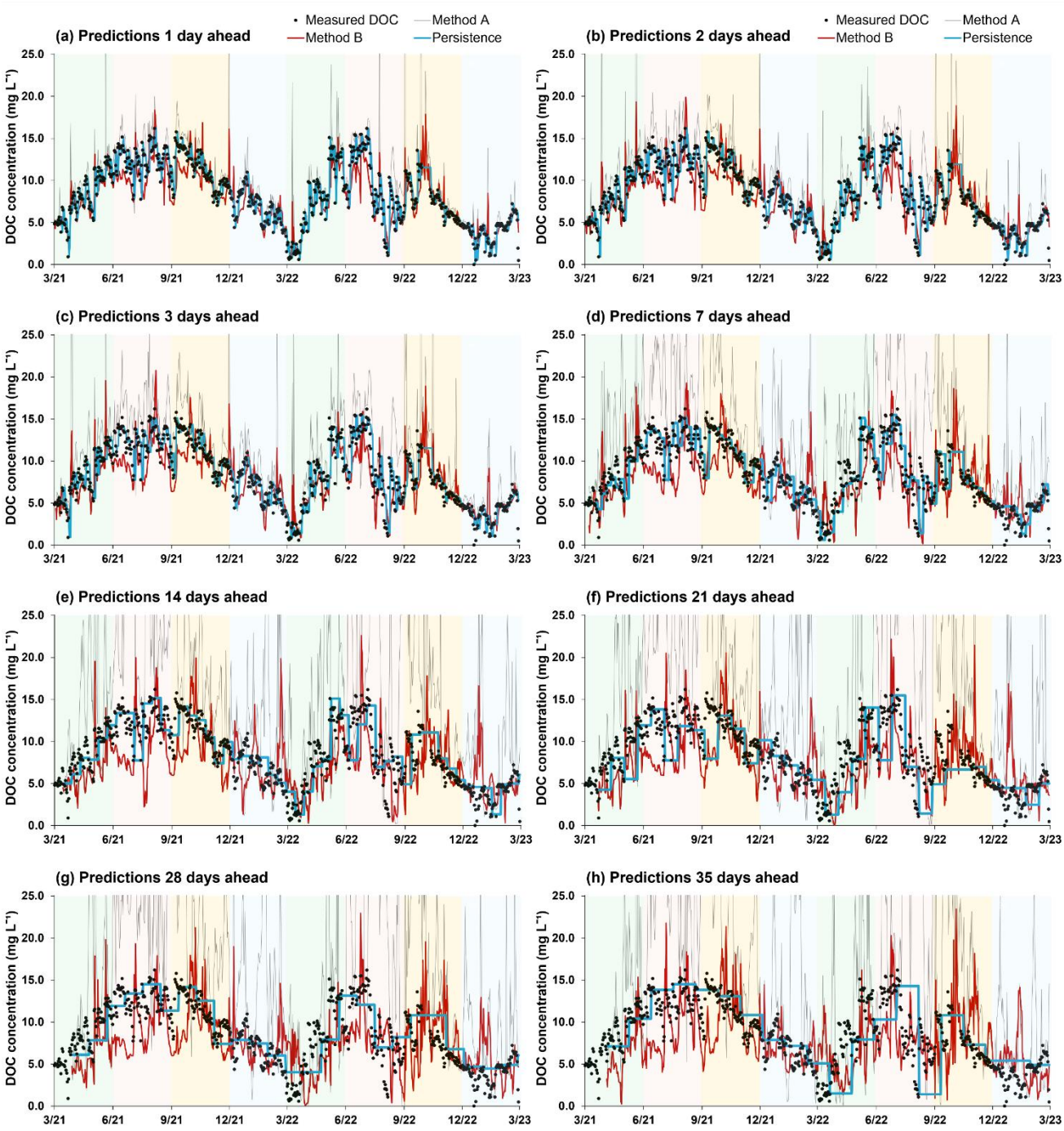


Figure 17. *DOC concentration forecasts produced using the process-based model and for Site Persistence for different forecast horizons (1-, 2-, 3-,7-, 14-, 21-, 28- and 35-days ahead) throughout the entire forecasting period plotted with the observed DOC concentrations used for evaluation. The background light shade colours in the graphs indicate spring (green), summer (red), autumn (yellow) and winter (blue) seasons, respectively. Method B indicates the successful run used for the evaluation of the forecasting, and method A indicates the worst run produced during workflow development which is displayed for reference.*

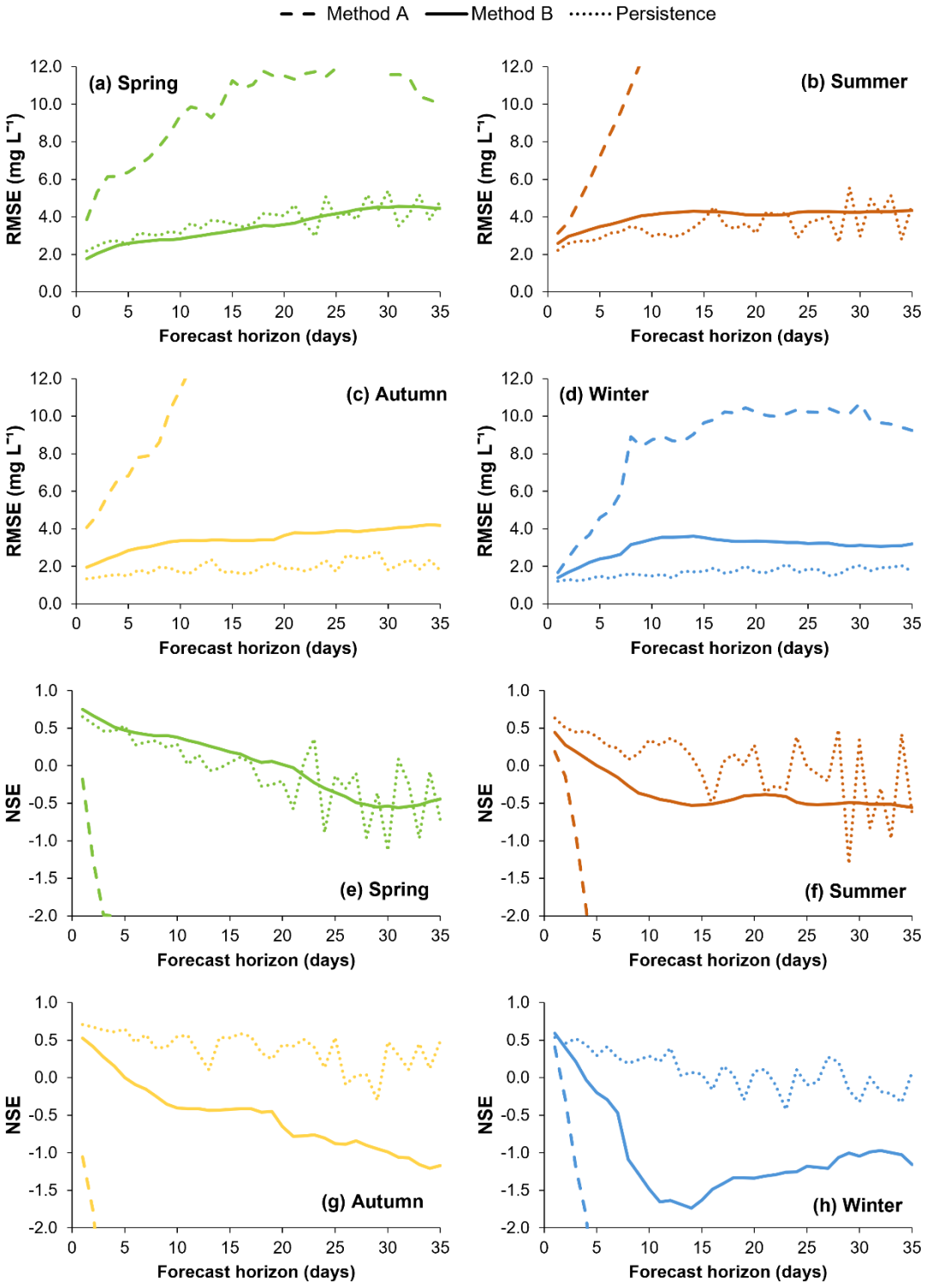


Figure 18. *Root mean squared error (RMSE) and Nash-Sutcliffe coefficient (NSE) performance for the process-based workflow and for Site Persistence for DOC concentration forecasts over the forecast horizon for the meteorological seasons in the forecasting period. Method B indicates the successful run used for the evaluation of the forecasting, and method A indicates the worst run produced during workflow development which is displayed for reference.*

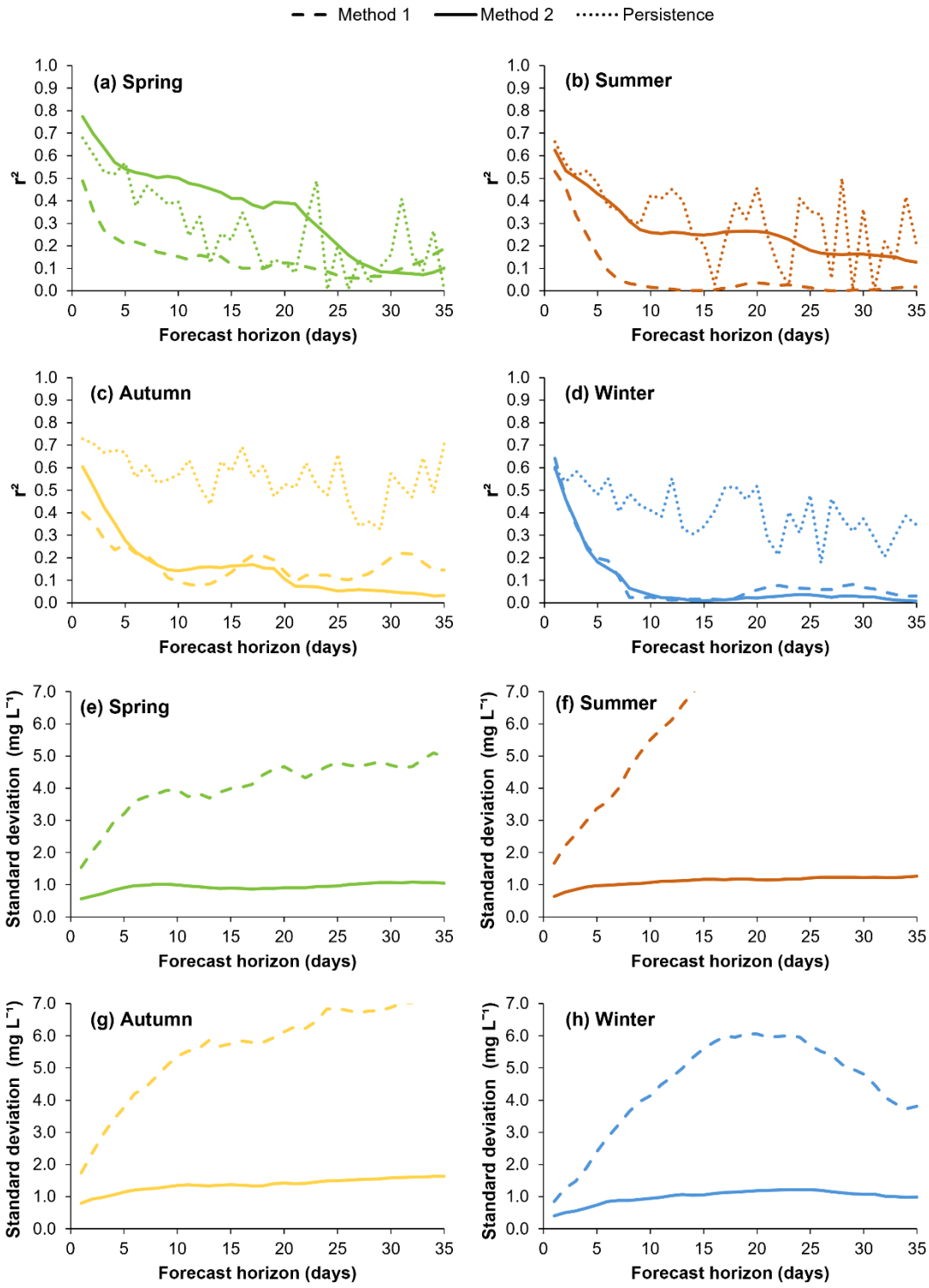


Figure 19. *Coefficient of determination (r^2) performance and quantified forecast uncertainty (standard deviation) of the process-based workflow andfor Site Persistence for DOC concentration forecasts over the forecast horizon for the meteorological seasons in the forecasting period. Method B indicates the successful run used for the evaluation of the forecasting, and method A indicates the worst run produced during workflow development which is displayed for reference.*

4. DISCUSSION

4.1 Scientific Insights

The most influential drivers for fDOM identified for each site in the seasonal workflow suggest potential carbon-producing processes that drive DOM dynamics in each lake that could reflect dynamics at their inflow river catchments as well. Interestingly, the most influential drivers were related to temperature and moisture at the deepest soil layers, indicating a common relevance for carbon inputs from terrestrial processes in both catchments. The Irish site is dominated by peat, a highly organic soil known to be sensitive to temperature-induced carbon release, especially as temperature levels rise. This is supported by the relation of soil temperatures with fDOM at this site (Ryder et al., 2014). In contrast, water availability constraints are much more pronounced in drier soils such as those of the Spanish site, as is validated by the partial dependence relationship of soil moisture with fDOM (Šimek et al., 2011). A plausible explanation is that, in Sau, increased soil moisture could boost biological activity, enhancing fDOM production, while in Feeagh, soil moisture showed a bell-shaped relationship, suggesting a more nuanced interplay between oxygen availability and microbial processes (Figure 6).

DOM dynamics are driven by physical and biogeochemical processes in the soil that are sensitive to changes in temperature and moisture, e.g., microbial processes that break down organic matter (Jennings et al., 2010; Kalbitz et al., 2000).The fact that the deepest layers of the soil were more important in the model for both sites than those shallower could be linked to potential carbon attenuation processes, such as soil organic matter decomposition and retention in the soil, including sorption processes (Jennings et al., 2020; Kritzberg et al., 2020; Klante et al., 2021). In any case, the production of carbon in the catchment that eventually ends up in the lake requires concurrent downstream transport, governed by rainfall events.

Climate and topography dictate the flushing of accumulated DOM during rainfall events, but also can influence sustained baseflow DOM contributions (Jennings et al., 2020; Anderson et al., 2023). In Feeagh, carbon exports from the catchment have been observed regularly throughout the entire annual cycle, with a seasonal variability (Doyle et al., 2019). In contrast, DOM in Sau accumulates primarily during the summer and is mainly flushed out via surface runoff during the wetter winter months (Marcé et al., 2021).

These patterns are supported by the relationship between inflow DOC concentration and fDOM at both sites, which shows a slight increase in predicted fDOM under lower carbon input conditions. Thus, inflow DOC concentration could reflect discharge pulses and dilution effects driven by precipitation (Jennings et al., 2020), following the characteristic seasonality of each site.

Seasonality plays a crucial role in fDOM predictions, as evidenced by the relationship of the Julian day driver with DOM dynamics at both sites. At the Irish site, DOM seasonality is primarily shaped by natural environmental processes, whereas in the Spanish site, human influence plays a much greater role. This distinction helps explain why surface lake water temperature and solar radiation, two variables typically linked to strong seasonal patterns, were important only for Feeagh, while reservoir volume was significant only for Sau. Volume and soil variables produce a similar effect on fDOM as Julian day at Sau, given that higher volumes closer to the winter season can lead to higher fDOM values. Incorporating Julian day into the workflow offers a simple yet effective way to represent seasonality, potentially replacing seasonal variables (e.g., air temperature) (see the correlation matrix of all drivers in Supplementary Material). This proves that the use of machine learning approaches opens up opportunities to assess diverse drivers under contrasting conditions.

4.2 Accuracy of the workflows

The seasonal-scale workflow demonstrated strong and consistent performance across both lakes, even when driven by a reduced set of reanalysis-based predictors. The Random Forest seasonal predictions outperformed climatology across the forecasting range, indicating that the workflow provides clear added value for users who require actionable information at seasonal lead times. The high correlation between seasonal forecast input variables and ERA5 observations, together with positive CRPSS values, further confirmed that the drivers themselves contain skill beyond statistical averages. This gave the resulting fDOM forecasts a strong foundation.

At the daily scale, the process-based forecasting workflow produced its highest accuracy during the spring, a period characterised by the highest natural variability in DOC concentrations. This is an important finding, as it indicates that the workflow adds value

precisely when DOC changes most rapidly and when operational decisions are most sensitive to short-term fluctuations. During summer and early autumn, forecast performance remained reasonable for short lead times, though less consistent than in spring. In contrast, predictive skill declined in late autumn and winter, when high discharge and less accurate precipitation forecasts reduced the capacity of the model to track fast hydrological fluctuations. Even under these conditions, however, the workflow produced forecasts with reliable short-horizon accuracy, and these horizons remain relevant for operational activities that benefit from 1–3 day anticipation.

The two modelling approaches demonstrate that different forecasting horizons can be served effectively: seasonal predictions deliver broad strategic guidance, while daily forecasts can provide short-notice information during periods of high vulnerability.

4.3 Scalability

Both workflows were designed to favour reproducibility and deployment beyond the study sites. In the lakes, the ability to maintain predictive performance using only globally available reanalysis variables demonstrates strong transfer potential. The fact that models remained robust even when influential site-specific variables were removed highlights that the essential processes governing fDOM dynamics are captured by variables that can be widely accessible. This enables the seasonal workflow to be applied in systems with limited monitoring infrastructure, provided that fDOM or a suitable DOM proxy is available for training.

The daily-scale workflow also shows potential for replication in catchments where a DOC process-based model can be calibrated using long-term meteorological and water quality observations. Only the initial carbon pool conditions were recalibrated during forecast generation, which reduced computational demands while preserving the integrity of the underlying model structure. This approach supports broader scalability, especially in contexts where DOC datasets are now increasingly collected through high-frequency sensors or derived proxies (Lofton et al., 2023). When combined, the two workflows offer a complementary forecasting strategy that can be extended across a range of hydro-climatic settings: machine learning for strategic seasonal predictions and process-based forecasting for short-term operational decision support.

4.4 Limitations and future research

Our approach offers the opportunity to validate and deploy workflows capable of delivering DOM predictions in both undisturbed and anthropized sites, even when only limited data on input drivers are available, while at the same time providing insights into the dominant drivers. However, a site-specific model validation, including identification of appropriate training and testing periods, hyperparameter tuning for each specific study site and assessment of overfitting is essential.

While the workflows performed relatively well, several limitations warrant consideration. In the lakes, the relationships revealed through machine learning do not always map directly onto mechanistic processes, even when they are consistent with known carbon-cycling behaviour. Interpretability tools such as partial dependence and SHAP alleviate this limitation, but future work should explore hybrid modelling frameworks. In the studied river, the daily-scale forecasts were limited by uncertainties in precipitation forecasts and by the structural constraints of the DOC model. The weather ensemble tended to underestimate rainfall extremes in autumn and winter, which reduced the accuracy of discharge estimates during high-flow periods. Since DOC washout is highly sensitive to such events, forecast skill naturally declined. Incorporating precipitation downscaling or local bias-correction methods would likely improve hydrological fidelity and, in turn, DOC predictions. Additionally, the use of a single data assimilation algorithm may have constrained uncertainty propagation. Exploring alternative assimilation techniques and frequencies or ensemble-based assimilation frameworks could help address the overconfidence identified in uncertainty estimates.

In terms of driver attribution, it is of note that the relationships identified using machine learning may not always be related at a process level (Toming et al., 2020; Sullivan, 2022). In our case, however, many of these same drivers had already been identified for river DOC levels in the Feeagh catchment including the Glenamong River (e.g., Doyle et al., 2019; Ryder et al., 2014). While the workflows can be easily replicated, fDOM data or data for another proxy for DOM are required. It is of note that such proxies of DOM are increasingly being incorporated into water quality monitoring programmes, an aspect that is convenient for testing workflows such as the one described (Downing et al., 2012).

Future research could expand the application of this framework to a broader range of lakes and/or rivers, integrate additional drivers such as remote sensing-derived terrestrial and aquatic quantity and quality parameters (Duan et al., 2025), explore hybrid modelling approaches, and deep learning.

5. CONCLUSIONS

This deliverable demonstrates that dissolved organic matter in source waters can be effectively predicted using a combination of data-driven and process-based modelling approaches tailored to different temporal scales. At the seasonal scale, the machine-learning workflow successfully provided insights into the environmental drivers most critical to DOM dynamics in contrasting hydroclimatic settings, with deep soil temperature dominating in the peat-rich temperate catchment and deep soil moisture in the drier Mediterranean system. The workflow retained strong predictive skill even when driven solely by open-access reanalysis and seasonal forecast data, and importantly, the resulting fDOM forecasts consistently outperformed climatology. This confirms their added value for utilities and water managers who require early-season guidance for planning and risk anticipation. By providing transparent driver attribution and maintaining robust performance with a minimal set of globally available predictors, the framework offers a scalable pathway for seasonal DOM forecasting in diverse freshwater environments.

At the daily scale, the process-based forecasting workflow showed that short-term DOC predictions in source waters can achieve meaningful accuracy, particularly during spring, the period when DOC concentrations can be, as in our study site, naturally most variable and likely to be operationally most challenging for treatment in other settings. Under these conditions, the workflow produced reliable forecasts several days ahead, demonstrating practical value for process optimisation, early-warning applications and near-term management decisions. Although performance declined in autumn and winter, the ability to anticipate DOC fluctuations during peak variability highlights the workflow's relevance for operational settings. Both the seasonal and daily forecasting components of this deliverable offer complementary tools for anticipating DOM dynamics across timescales, advancing the capacity to support source-water protection, treatment planning and proactive decision-making.

REFERENCES

Anderson, L.E., DeMont, I., Dunnington, D.D., Bjorndahl, P., Redden, D.J., Brophy, M.J., Gagnon, G.A., 2023. A review of long-term change in surface water natural organic matter concentration in the northern hemisphere and the implications for drinking water treatment. Science of The Total Environment 858, 159699. https://doi.org/10.1016/j.scitotenv.2022.159699

Arvola, L., George, G., Livingstone, D.M., Järvinen, M., Blenckner, T., Dokulil, M.T., Jennings, E., Aonghusa, C.N., Nõges, P., Nõges, T., Weyhenmeyer, G.A., 2010. The Impact of the Changing Climate on the Thermal Characteristics of Lakes, in: George, G. (Ed.), The Impact of Climate Change on European Lakes. Springer Netherlands, Dordrecht, pp. 85–101. https://doi.org/10.1007/978-90-481-2945-4_6

Asadollah, S.B.H.S., Safaeinia, A., Jarahizadeh, S., Alcalá, F.J., Sharafati, A., Jodar-Abellan, A., 2025. Dissolved organic carbon estimation in lakes: Improving machine learning with data augmentation on fusion of multi-sensor remote sensing observations. Water Research 277, 123350. https://doi.org/10.1016/j.watres.2025.123350

Beven, K.J., 2009. Environmental modelling : an uncertain future? : an introduction to techniques for uncertainty estimation in environmental prediction / Keith Beven. Routledge, London ;

Bradford, J.B., Weltzin, J., McCormick, M.L., Baron, J., Bowen, Z., Bristol, S., Carlisle, D., Crimmins, T., Cross, P.C., DeVivo, J., Dietze, M., Freeman, M., Goldberg, J., Hooten, M., Hsu, L., Jenni, K., Keisman, J.L., Kennen, J., Lee, K., Lesmes, D.P., Loftin, K., Miller, B.W., Murdoch, P.S., Newman, J., Prentice, K.L., Rangwala, I., Read, J., Sieracki, J., Sofaer, H., Thur, S., Toevs, G., Werner, F., White, C.L., White, T., Wiltermuth, M.T., 2020. Ecological forecasting—21st century science for 21st century management (Report No. 2020–1073), Open-File Report. Reston, VA. https://doi.org/10.3133/ofr20201073

Carey, C.C., Woelmer, W.M., Lofton, M.E., Figueiredo, R.J., Bookout, B.J., Corrigan, R.S., Daneshmand, V., Hounshell, A.G., Howard, D.W., Lewis, A.S.L., McClure, R.P., Wander, H.L., Ward, N.K., Thomas, R.Q., 2022. Advancing lake and reservoir water quality management with near-term, iterative ecological forecasting. Inland Waters 12, 107–120. https://doi.org/10.1080/20442041.2020.1816421

Creed, I.F., Bergström, A.-K., Trick, C.G., Grimm, N.B., Hessen, D.O., Karlsson, J., Kidd, K.A., Kritzberg, E., McKnight, D.M., Freeman, E.C., Senar, O.E., Andersson, A., Ask, J., Berggren, M., Cherif, M., Giesler, R., Hotchkiss, E.R., Kortelainen, P., Palta, M.M., Vrede, T., Weyhenmeyer, G.A., 2018. Global change-driven effects on dissolved organic matter composition: Implications for food webs of northern lakes. Global Change Biology 24, 3692–3714. https://doi.org/10.1111/gcb.14129

Dietze, M.C., 2017. Ecological Forecasting. Princeton University Press. https://doi.org/10.2307/j.ctvc7796h

Downing, B.D., Pellerin, B.A., Bergamaschi, B.A., Saraceno, J.F., Kraus, T.E.C., 2012. Seeing the light: The effects of particles, dissolved materials, and temperature on in situ measurements of

DOM fluorescence in rivers and streams. Limnology and Oceanography: Methods 10, 767–775. https://doi.org/10.4319/lom.2012.10.767

Doyle, B.C., de Eyto, E., Dillane, M., Poole, R., McCarthy, V., Ryder, E., Jennings, E., 2019. Synchrony in catchment stream colour levels is driven by both local and regional climate. Biogeosciences 16, 1053–1071. https://doi.org/10.5194/bg-16-1053-2019

Doyle, B.C., De Eyto, E., McCarthy, V., Dillane, M., Jennings, E., 2023. The organic carbon budget of an oligotrophic temperate peatland lake. Limnology & Oceanography 68, 544–556. https://doi.org/10.1002/lno.12292

Duan, H., Cao, Z., Luo, J., Shen, M., 2025. AI-driven opportunities and challenges in lake remote sensing. Information Geography 1, 100014. https://doi.org/10.1016/j.infgeo.2025.100014

Harkort, L., Duan, Z., 2023. Estimation of dissolved organic carbon from inland waters at a large scale using satellite data and machine learning methods. Water Research 229, 119478. https://doi.org/10.1016/j.watres.2022.119478

Hipsey, M.R., Bruce, L.C., Boon, C., Busch, B., Carey, C.C., Hamilton, D.P., Hanson, P.C., Read, J.S., de Sousa, E., Weber, M., Winslow, L.A., 2019. A General Lake Model (GLM 3.0) for linking with high-frequency sensor data from the Global Lake Ecological Observatory Network (GLEON). Geoscientific Model Development 12, 473–523. https://doi.org/10.5194/gmd-12-473-2019

Hollister, J.W., Milstead, W.B., Kreakie, B.J., 2016. Modeling lake trophic state: a random forest approach. Ecosphere 7, e01321. https://doi.org/10.1002/ecs2.1321

Jennings, E., de Eyto, E., Moore, T., Dillane, M., Ryder, E., Allott, N., Nic Aonghusa, C., Rouen, M., Poole, R., Pierson, D.C., 2020. From Highs to Lows: Changes in Dissolved Organic Carbon in a Peatland Catchment and Lake Following Extreme Flow Events. Water 12. https://doi.org/10.3390/w12102843

Jennings, E., Järvinen, M., Allott, N., Arvola, L., Moore, K., Naden, P., Aonghusa, C.N., Nõges, T., Weyhenmeyer, G.A., 2010. Impacts of Climate on the Flux of Dissolved Organic Carbon from Catchments, in: George, G. (Ed.), The Impact of Climate Change on European Lakes. Springer Netherlands, Dordrecht, pp. 199–220. https://doi.org/10.1007/978-90-481-2945-4_12

Kalbitz, K., Solinger, S., Park, J.-H., Michalzik, B., Matzner, E., 2000. CONTROLS ON THE DYNAMICS OF DISSOLVED ORGANIC MATTER IN SOILS: A REVIEW. Soil Science 165, 277.

Klante, C., Larson, M., Persson, K.M., 2021. Brownification in Lake Bolmen, Sweden, and its relationship to natural and human-induced changes. Journal of Hydrology: Regional Studies 36, 100863. https://doi.org/10.1016/j.ejrh.2021.100863

Kritzberg, E.S., Hasselquist, E.M., Škerlep, M., Löfgren, S., Olsson, O., Stadmark, J., Valinia, S., Hansson, L.-A., Laudon, H., 2020. Browning of freshwaters: Consequences to ecosystem services, underlying drivers, and potential mitigation measures. Ambio 49, 375–390. https://doi.org/10.1007/s13280-019-01227-5

Lake, P.S., Palmer, M.A., Biro, P., Cole, J., Covich, A.P., Dahm, C., Gibert, J., Goedkoop, W., Martens, K., Verhoeven, J., 2000. Global Change and the Biodiversity of Freshwater Ecosystems: Impacts on Linkages between Above-Sediment and Sediment Biota: All forms of anthropogenic disturbance—changes in land use, biogeochemical processes, or biotic addition or loss—not only

damage the biota of freshwater sediments but also disrupt the linkages between above-sediment and sediment-dwelling biota. BioScience 50, 1099–1107. https://doi.org/10.1641/0006-3568(2000)050%255B1099:GCATBO%255D2.0.CO;2

Lofton, M.E., Howard, D.W., Thomas, R.Q., Carey, C.C., 2023. Progress and opportunities in advancing near-term forecasting of freshwater quality. Global Change Biology 29, 1691–1714. https://doi.org/10.1111/gcb.16590

Marcé, R., Verdura, L., Leung, N., 2021. Dissolved organic matter spectroscopy reveals a hot spot of organic matter changes at the river–reservoir boundary. Aquat Sci 83, 67. https://doi.org/10.1007/s00027-021-00823-6

Mercado-Bettín, D., Clayer, F., Shikhani, M., Moore, T.N., Frías, M.D., Jackson-Blake, L., Sample, J., Iturbide, M., Herrera, S., French, A.S., Norling, M.D., Rinke, K., Marcé, R., 2021. Forecasting water temperature in lakes and reservoirs using seasonal climate prediction. Water Research 201, 117286. https://doi.org/10.1016/j.watres.2021.117286

Meyer-Jacob, C., Michelutti, N., Paterson, A.M., Cumming, B.F., Keller, W. (Bill), Smol, J.P., 2019. The browning and re-browning of lakes: Divergent lake-water organic carbon trends linked to acid deposition and climate change. Sci Rep 9, 16676. https://doi.org/10.1038/s41598-019-52912-0

Murphy, A.H., 1992. Climatology, Persistence, and Their Linear Combination as Standards of Reference in Skill Scores. Weather and Forecasting 7, 692–698. https://doi.org/10.1175/1520-0434(1992)007%253C0692:CPATLC%253E2.0.CO;2

Paíz, R., Pierson, D.C., Lindqvist, K., Naden, P.S., de Eyto, E., Dillane, M., McCarthy, V., Linnane, S., Jennings, E., 2025a. Accounting for model parameter uncertainty provides more robust projections of dissolved organic carbon dynamics to aid drinking water management. Water Research 123238. https://doi.org/10.1016/j.watres.2025.123238

Paíz, R., Thomas, R.Q., Carey, C.C., de Eyto, E., Jones, I.D., Delany, A.D., Poole, R., Nixon, P., Dillane, M., McCarthy, V., Linnane, S., Jennings, E., 2025b. Near-term lake water temperature forecasts can be used to anticipate the ecological dynamics of freshwater species. Ecosphere 16, e70335. https://doi.org/10.1002/ecs2.70335

Pedregal-Montes, A., Jennings, E., Kothawala, D., Jones, K., Sjöstedt, J., Langenheder, S., Marcé, R., Farré, M.J., 2024. Disinfection by-product formation potential in response to variability in dissolved organic matter and nutrient inputs: Insights from a mesocosm study. Water Research 258, 121791. https://doi.org/10.1016/j.watres.2024.121791

Piai, L., Mei, S., van Gijn, K., Langenhoff, A., 2024. Effects of organic matter in drinking water and wastewater on micropollutant adsorption to activated carbon. Int. J. Environ. Sci. Technol. 21, 2547–2558. https://doi.org/10.1007/s13762-023-05132-z

Regier, P., Duggan, M., Myers-Pigg, A., Ward, N., 2023. Effects of random forest modeling decisions on biogeochemical time series predictions. Limnology and Oceanography: Methods 21, 40–52. https://doi.org/10.1002/lom3.10523

Riyadh, A., Peleato, N.M., 2024. Natural Organic Matter Character in Drinking Water Distribution Systems: A Review of Impacts on Water Quality and Characterization Techniques. Water 16, 446. https://doi.org/10.3390/w16030446

Ryder, E., de Eyto, E., Dillane, M., Poole, R., Jennings, E., 2014. Identifying the role of environmental drivers in organic carbon export from a forested peat catchment. Science of The Total Environment 490, 28–36. https://doi.org/10.1016/j.scitotenv.2014.04.091

Šimek, K., Comerma, M., García, J.-C., Nedoma, J., Marcé, R., Armengol, J., 2011. The Effect of River Water Circulation on the Distribution and Functioning of Reservoir Microbial Communities as Determined by a Relative Distance Approach. Ecosystems 14, 1–14. https://doi.org/10.1007/s10021-010-9388-4

Sullivan, E., 2022. Understanding from Machine Learning Models. The British Journal for the Philosophy of Science 73, 109–133. https://doi.org/10.1093/bjps/axz035

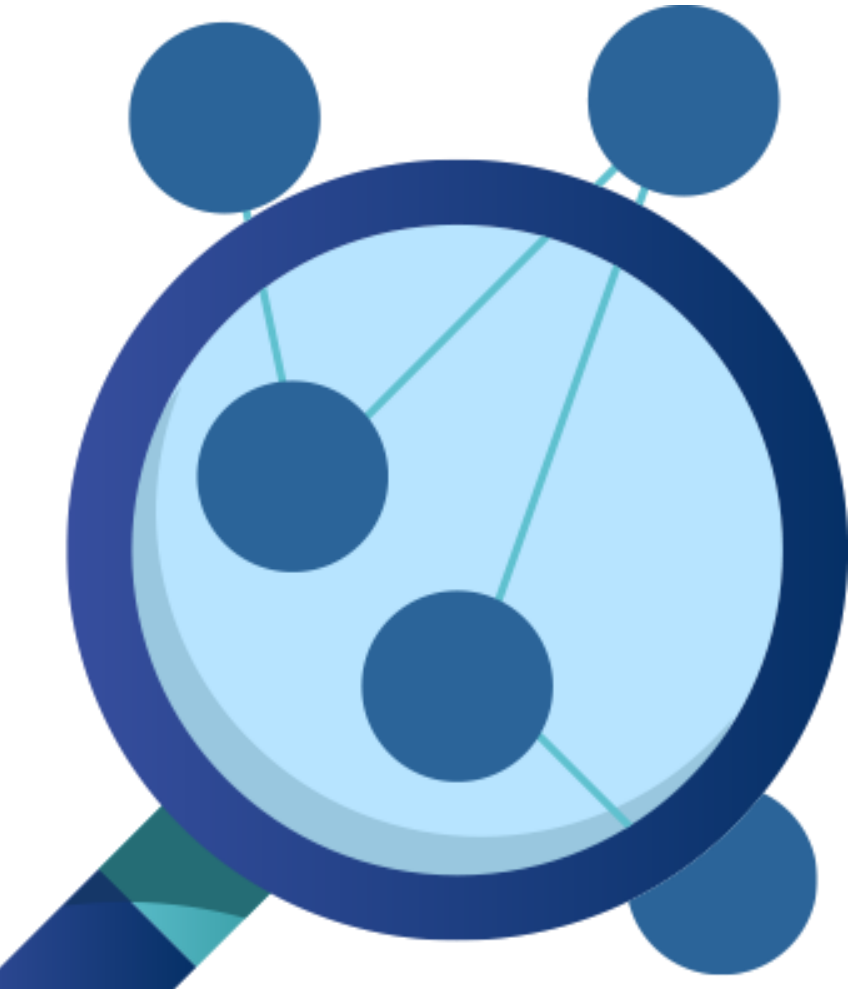
Thomas, R.Q., Figueiredo, R.J., Daneshmand, V., Bookout, B.J., Puckett, L.K., Carey, C.C., 2020. A Near-Term Iterative Forecasting System Successfully Predicts Reservoir Hydrodynamics and Partitions Uncertainty in Real Time. Water Resources Research 56, e2019WR026138. https://doi.org/10.1029/2019WR026138

Toming, K., Kotta, J., Uuemaa, E., Sobek, S., Kutser, T., Tranvik, L.J., 2020. Predicting lake dissolved organic carbon at a global scale. Sci Rep 10, 8471. https://doi.org/10.1038/s41598-020-65010-3

Xenopoulos, M.A., Barnes, R.T., Boodoo, K.S., Butman, D., Catalán, N., D'Amario, S.C., Fasching, C., Kothawala, D.N., Pisani, O., Solomon, C.T., Spencer, R.G.M., Williams, C.J., Wilson, H.F., 2021. How humans alter dissolved organic matter composition in freshwater: relevance for the Earth's biogeochemistry. Biogeochemistry 154, 323–348. https://doi.org/10.1007/s10533-021-00753-3

Xu, J., Morris, P.J., Liu, J., Ledesma, J.L.J., Holden, J., 2020. Increased Dissolved Organic Carbon Concentrations in Peat-Fed UK Water Supplies Under Future Climate and Sulfate Deposition Scenarios. Water Resources Research 56, e2019WR025592. https://doi.org/10.1029/2019WR025592

Zwart, J.A., Oliver, S.K., Watkins, W.D., Sadler, J.M., Appling, A.P., Corson-Dosch, H.R., Jia, X., Kumar, V., Read, J.S., 2023. Near-term forecasts of stream temperature using deep learning and data assimilation in support of management decisions. JAWRA Journal of the American Water Resources Association 59, 317–337. https://doi.org/10.1111/1752-1688.13093



intoDBP

Funded by
the European Union

SUPPLEMENTARY MATERIAL

fDOM correction for both study sites

In Feeagh, fDOM data was corrected for the temperature quenching effect in previous scientific studies (Doyle et al., 2019; Ryder et al., 2012).

In Sau, the fDOM data were corrected for the temperature quenching effect, following a test provided by the fDOM sensor manufacturer, where they use a 300 QSU sample of water and change the temperature to get the effect of temperature in the measurement, here are the results:

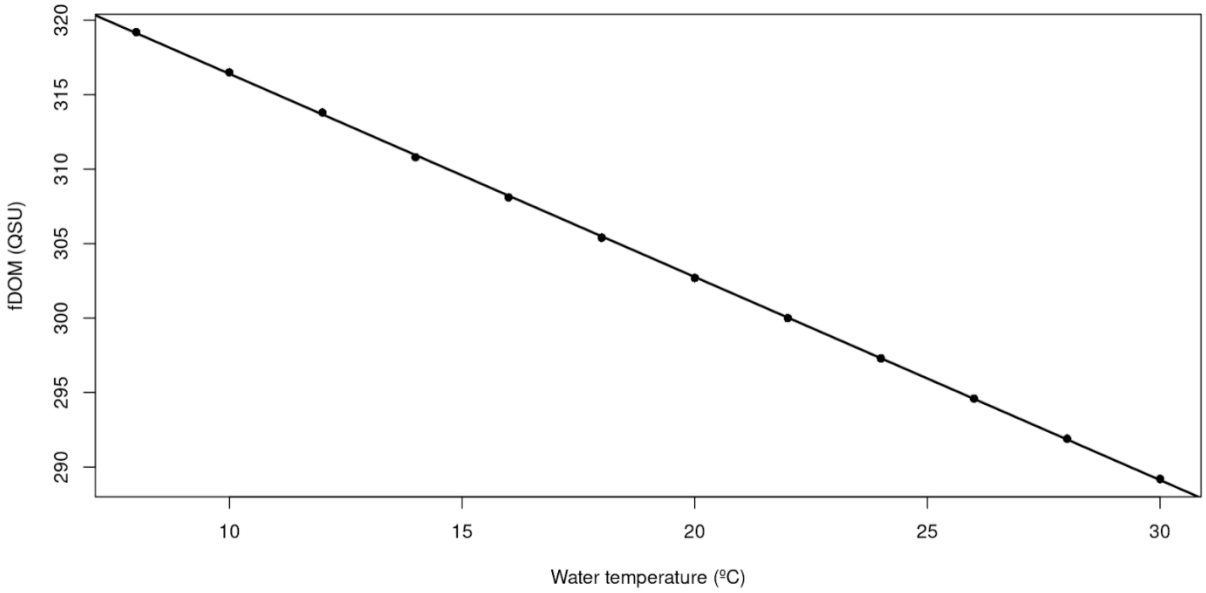


***Figure S1**. fDOM relation with temperature with a sample of 300 QSU provided by the manufacturer of the fDOM sensor in Sau.*

Then, fDOM data in Sau was corrected following this linear regression and surface water temperature on the lake. Here are the uncorrected values in blue and corrected values in red (8 negative values were removed from the total sample of 777):

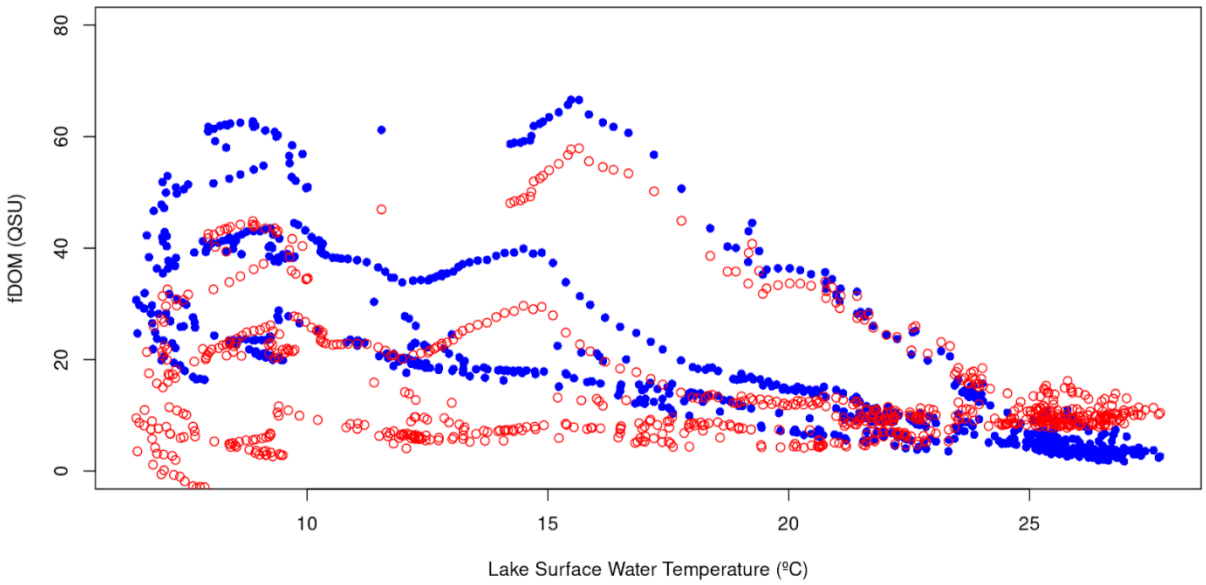


***Figure S2**. Uncorrected (blue) and corrected (red) fDOM data for Sau.*

Hydrologic Modelling

Daily time series of inflow discharge and inflow DOC concentration into each site were generated using the Generalised Watershed Loading Functions Model (GWLF) coupled with a DOC module (GWLF-DOC). This model simulates catchment hydrology (water balance and water distribution among the different hydrological pathways) and DOC dynamics (DOC production and DOC washout) in a daily time step. The model input requirements include daily time series of two meteorological variables: total precipitation and air temperature; as well as land cover, land use, and soil characterisation.

GWLF-DOC was applied to Feeagh based on previous model applications in the Irish catchment (Paíz et al., 2025), for which measured discharge data were used to calibrate and validate the hydrology (2013-2018 and 2019-2023, respectively), and DOC concentration data were used to calibrate the DOC module (2016-2023). In Sau, observed inflow discharge data were used to calibrate and validate the hydrology (2008-2011 and 2011-2024, respectively), and measured DOC concentration data were used to calibrate and validate the DOC module (2008-2014 and 2016-2018, respectively) using the same calibration strategy than for the Irish site. Calibration results were satisfactory for both hydrology (Feeagh: R^2 = 0.64 and NSE = 0.64; Sau: R^2 = 0.66 and NSE = 0.66;) and DOC (Feeagh: R^2 = 0.45 and NSE = 0.47; Sau: R^2 = 0.44 and NSE = 0.40). Similarly, validation results were satisfactory for both hydrology (Feeagh: R^2 = 0.60 and NSE = 0.60; Sau: R^2 = 0.42 and NSE = 0.42) and DOC in the case of Sau (R^2 = 0.50 and NSE = 0.46).

Lake Modelling

Daily time series of 5 key lake variables (see Fig. 2) were obtained from the General Lake Model (GLM) run for each site. GLM is an open-source, one-dimensional hydrodynamic model designed to simulate the vertical stratification and water balance of lakes and reservoirs. It calculates vertical profiles of temperature, and density by accounting for factors such as inflows and outflows, mixing processes, and surface heating and cooling (Hipsey et al., 2019).

GLM was calibrated and validated by evaluating the fit of modelled water temperature against measured water temperature profile data in Feeagh (2010-2015 and 2016-2017, respectively) and Sau (1997-2007 and 2008-2018, respectively). The calibration strategy was based on previous lake modelling deployments at each site (Mercado-Bettín et al., 2021; Paíz et al., 2024). Model performance was satisfactory for both sites.

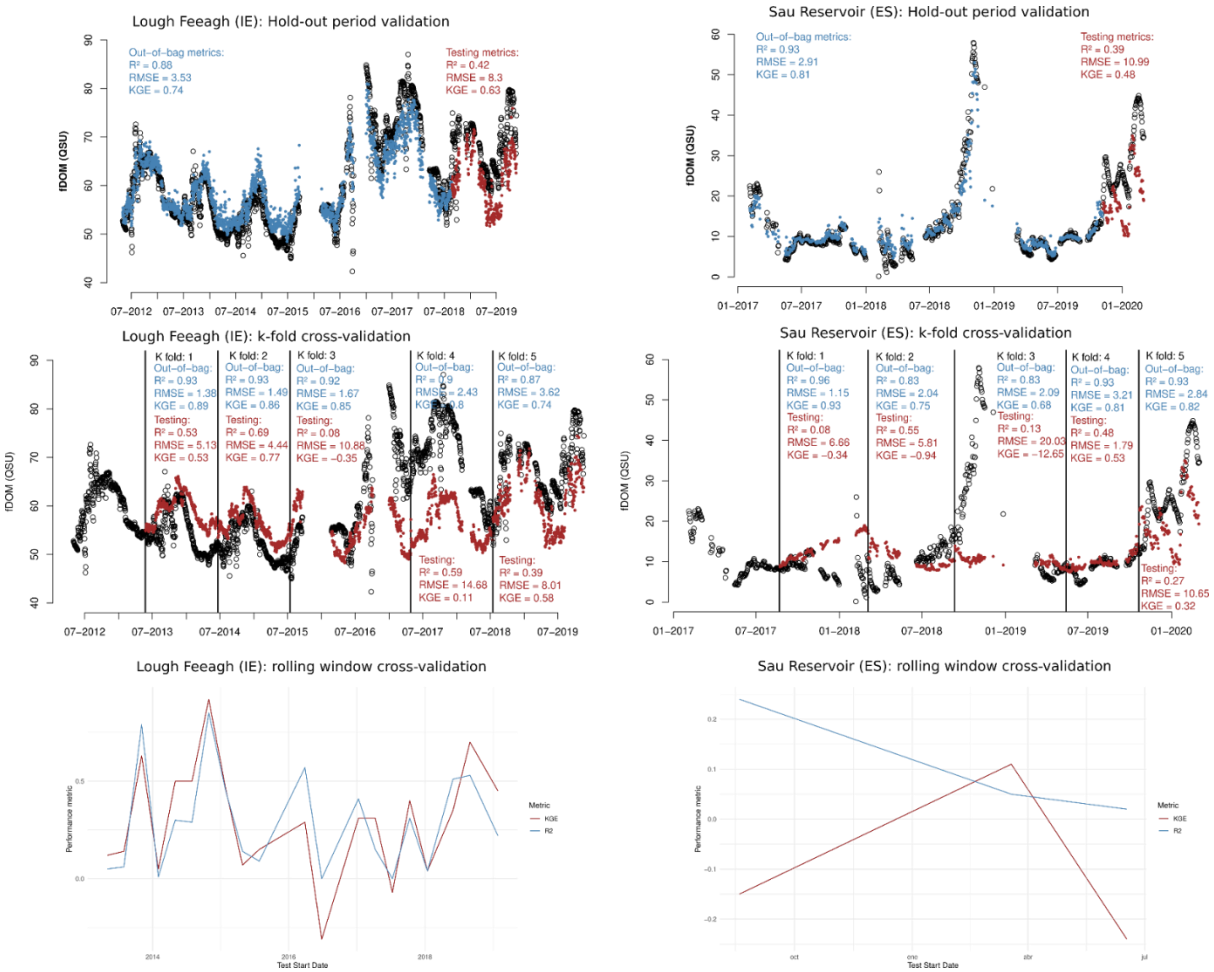


Figure S3. *Comparison of validation methods using random forest for both study sites: hold-out period method used in the main manuscript, k-fold cross-validation method using k=5, and rolling window cross-validation using training size of two year, testing size of one year, and a shift window every 90 days. For this method, in the case of Sau Reservoir it is not possible to get a clear analysis due to the limited data.*

Table S1. *Model comparison during training to predict fDOM using all selected drivers in both study sites. Statistic metrics (R^2, RMSE and KGE) were calculated to compare the performance of the models Random Forest (RF), eXtreme Gradient Boosting (XGBoost), Light Gradient Boosting (LGB), Catboosting (CTB), k-Nearest Neighbors (KNN), Support Vector Regression (SVR) and linear model during training period.*

Lough Feeagh								**Sau Reservoir**						
Model/Metric	RF	XGB	LGB	CTB	LM	KNN	SVR	RF	XGB	LGB	CTB	LM	KNN	SVR
R2	0.99	1	0.88	0.71	0.22	0.99	0.89	0.99	1	0.98	0.99	0.67	0.98	0.8
RMSE	0.97	0.4	3.35	5.19	7.87	0.92	2.95	0.81	0.11	1.48	0.72	5.43	1.17	1.38
KGE	0.95	0.99	0.79	0.59	0.25	0.98	0.9	0.97	1	0.97	0.99	0.75	0.99	0.98

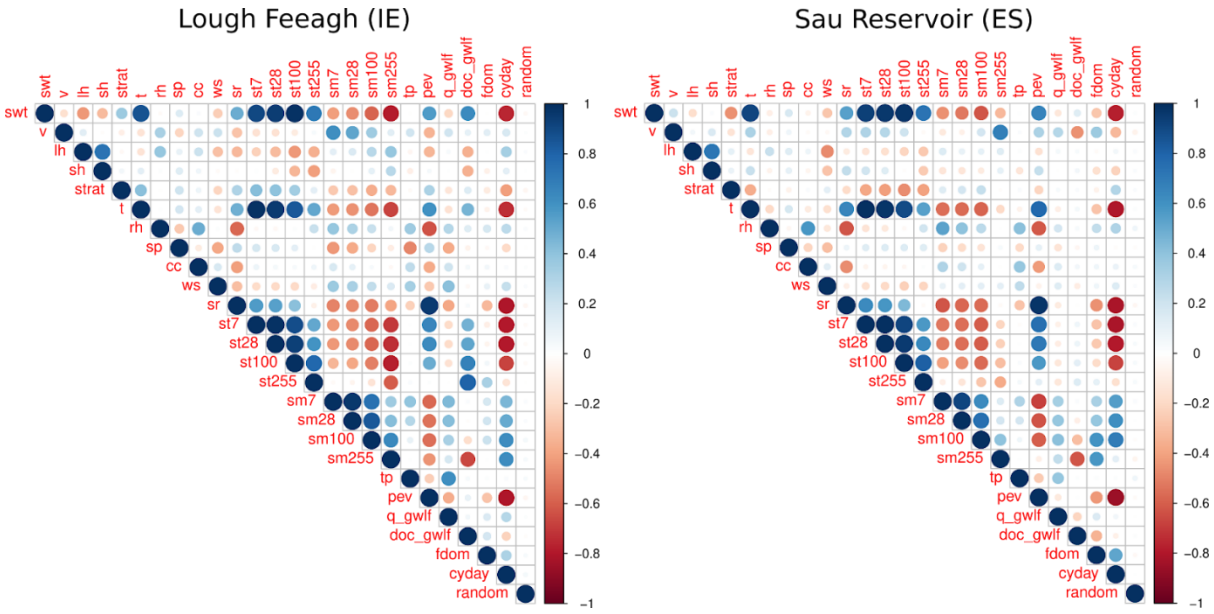
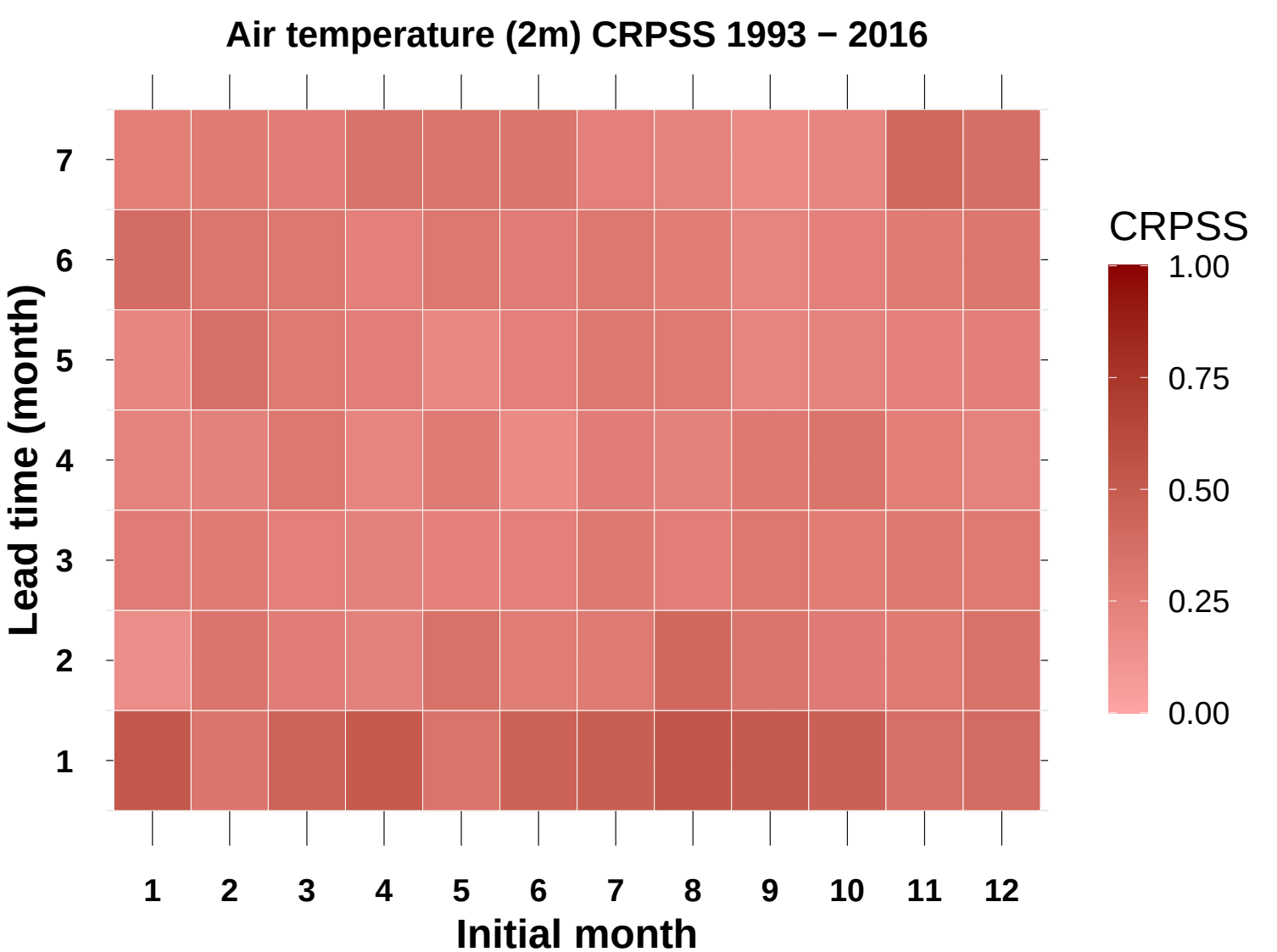
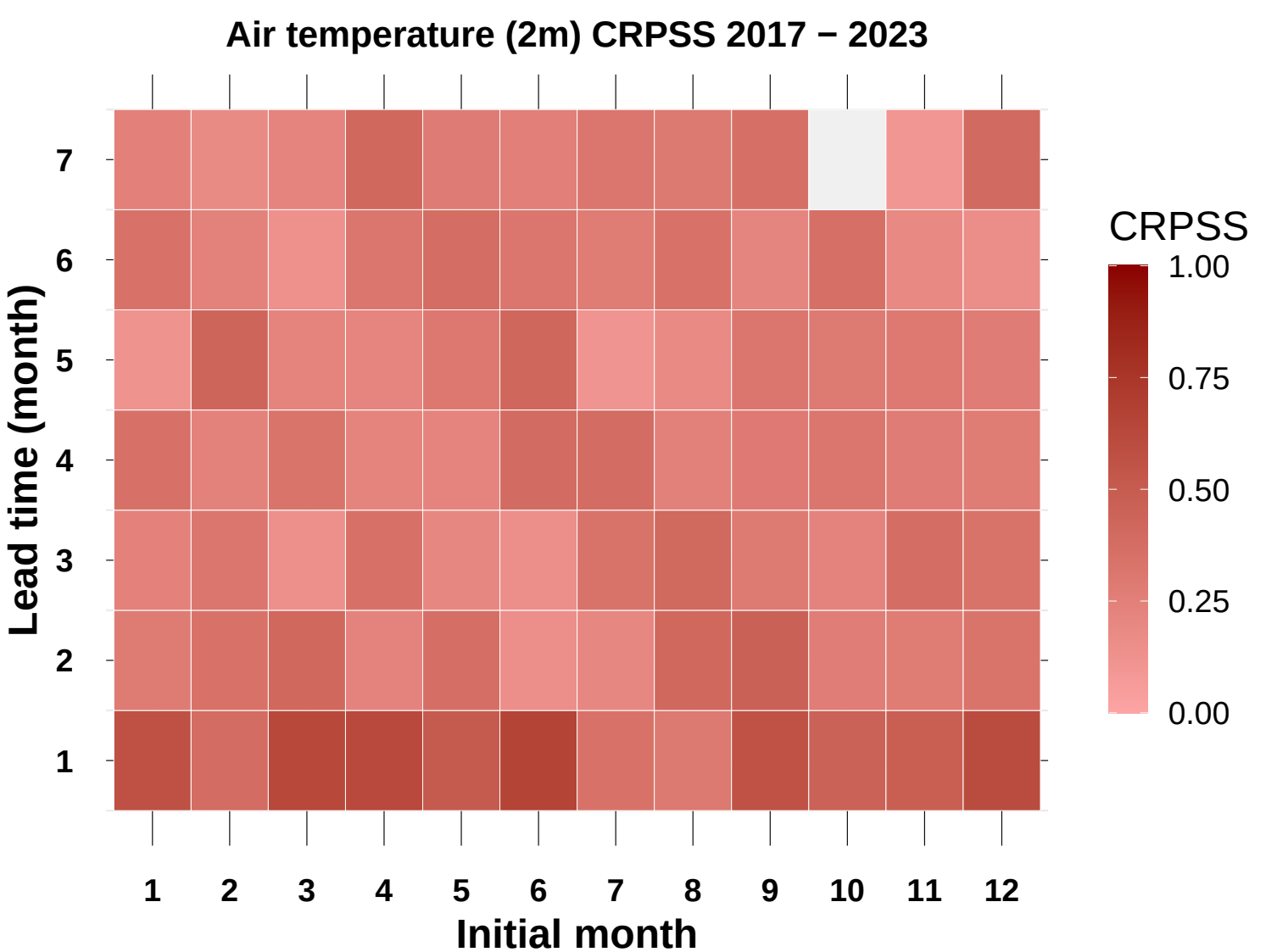
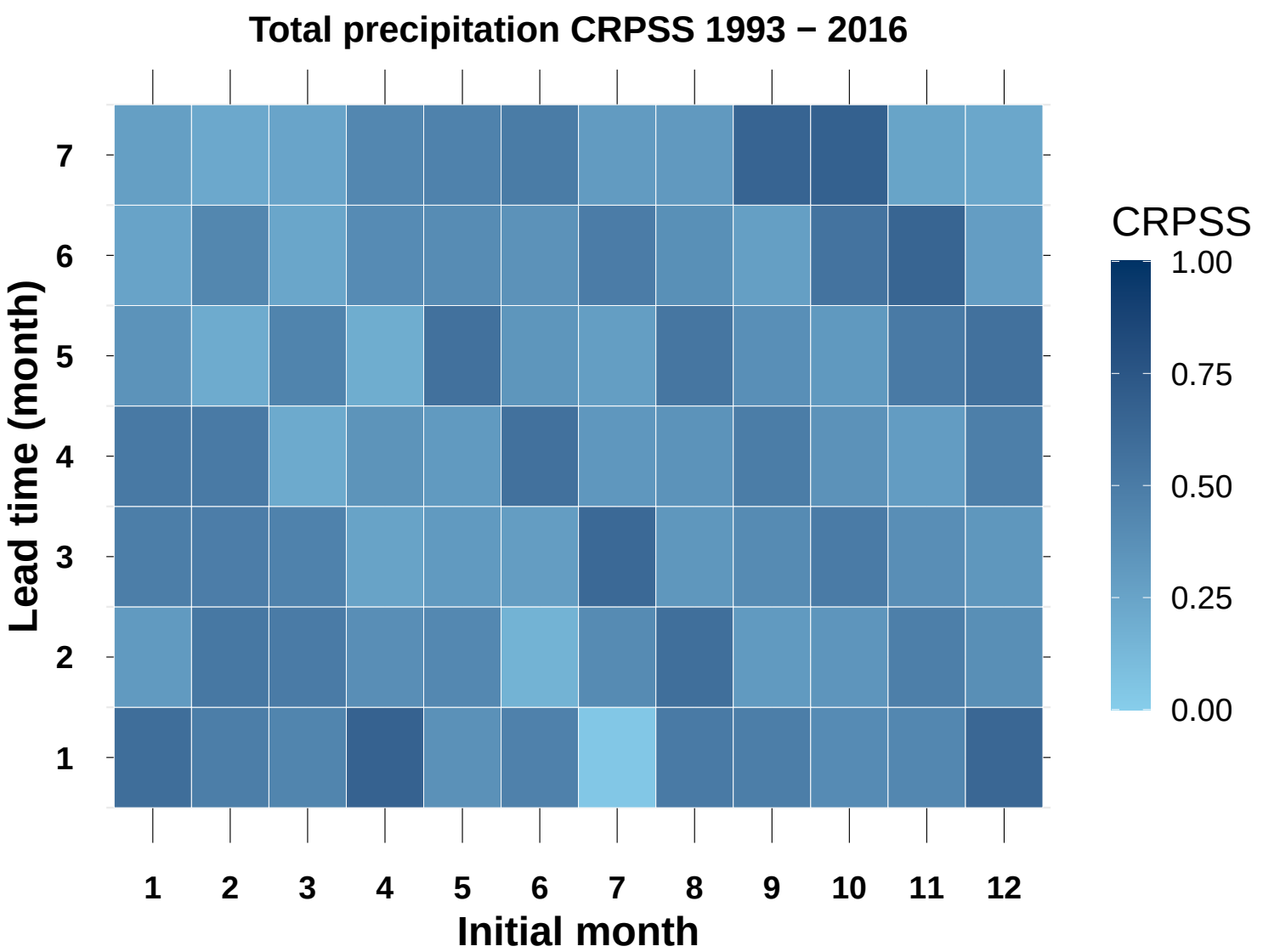


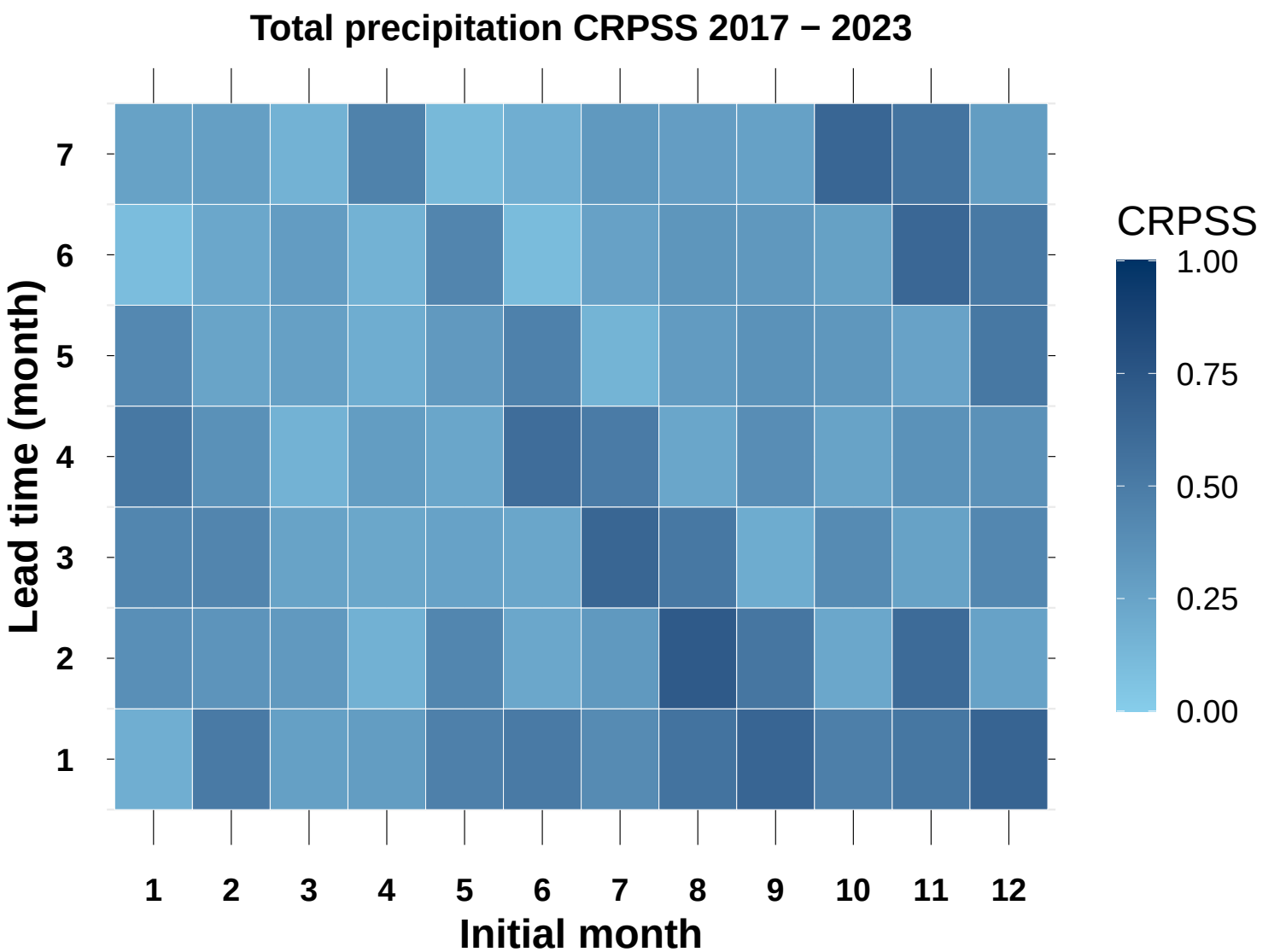
Figure S4. *Correlation matrix of all drivers and fDOM for Feeagh and Sau.*

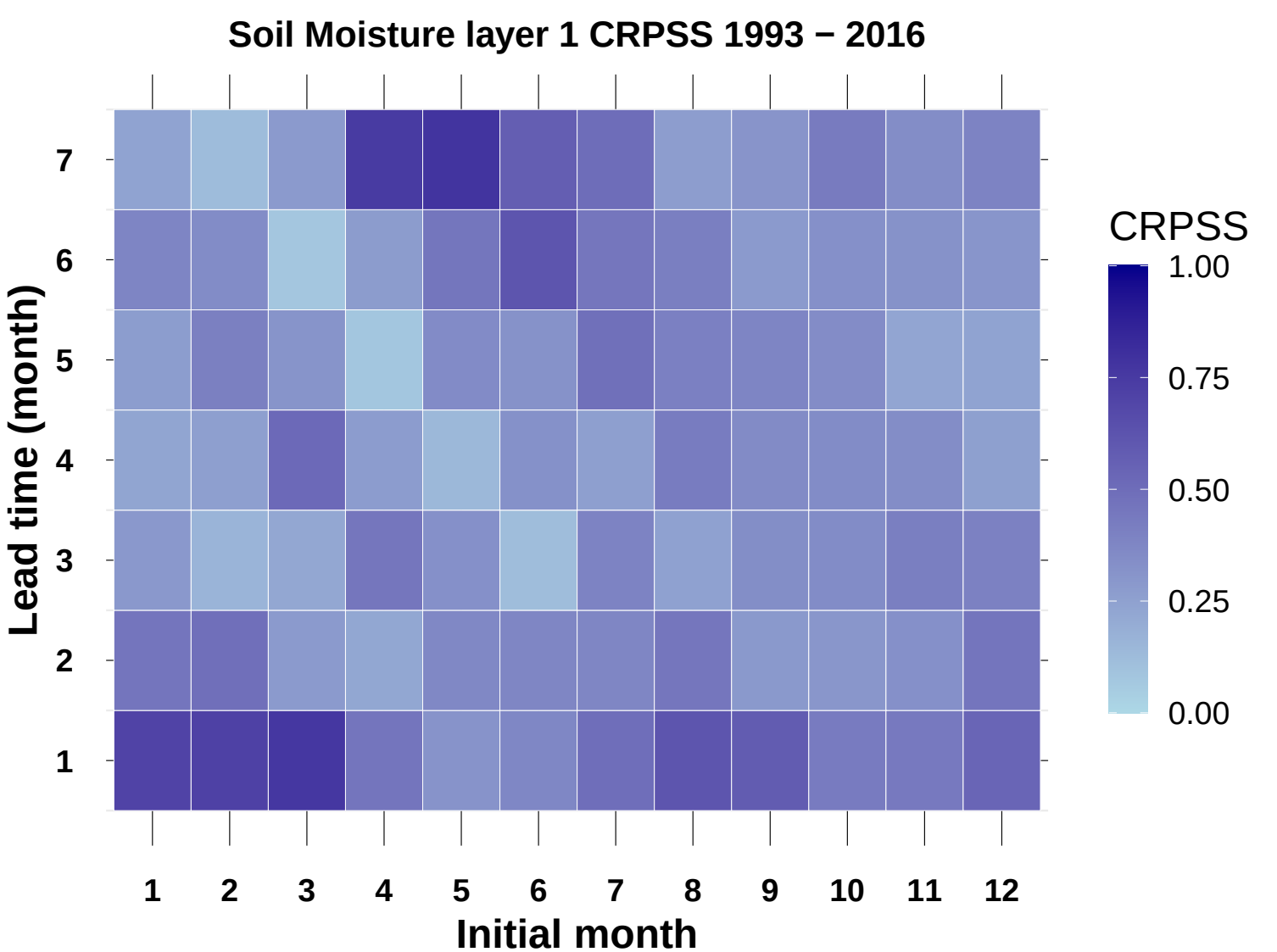
The following figures correspond to the verification of the seasonal forecast product against ERA5 observations for the meteorological and soil variables used to drive the seasonal fDOM forecasts in the seasonal workflow. **The results for Lough Feeagh are presented first, followed by the corresponding results for Sau Reservoir.** These heat maps show the performance of the seasonal forecast input data relative to observed conditions over the evaluation periods, providing context for the forecast skill assessments presented in the main document.

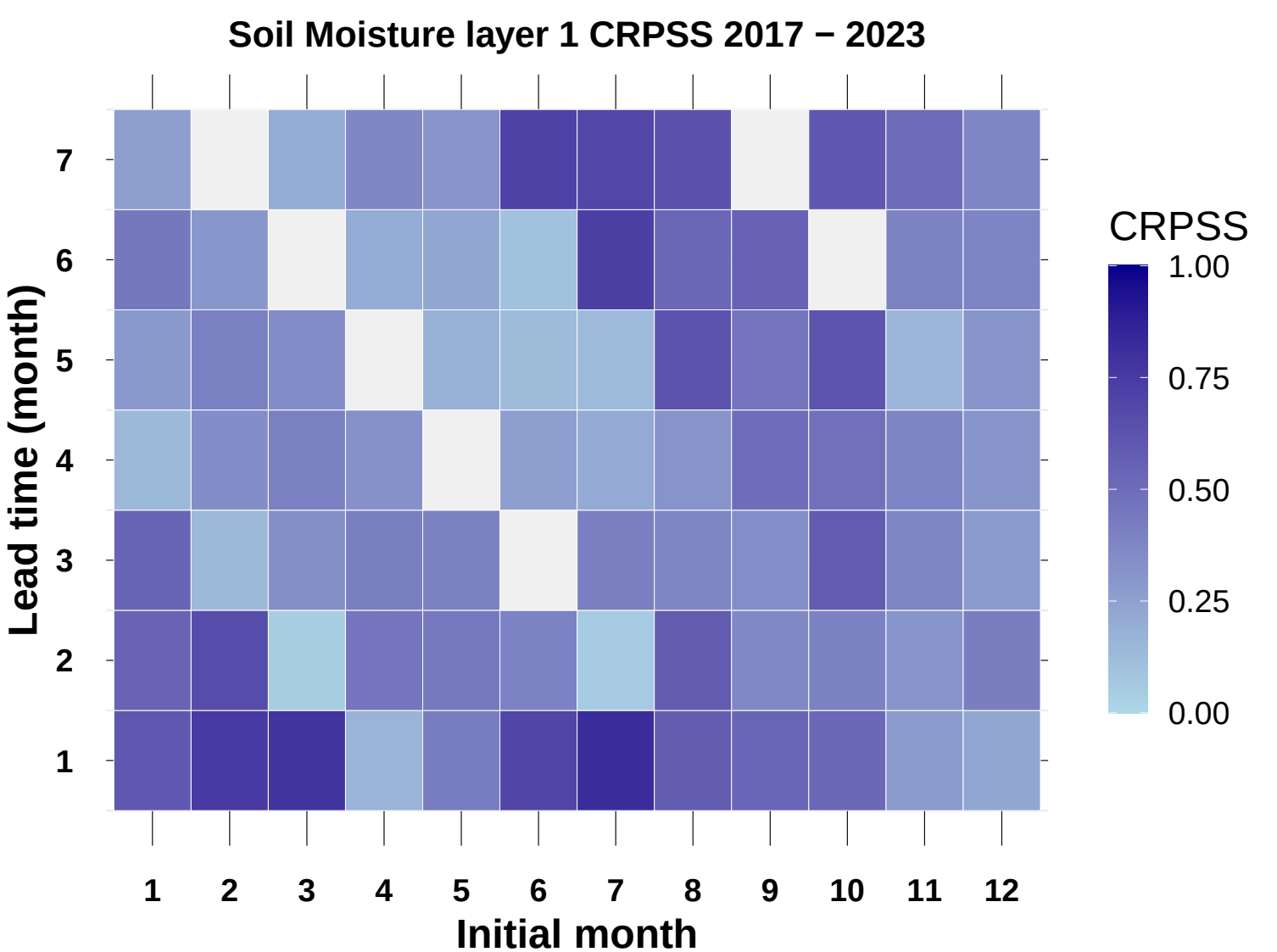

Air temperature (2m) CRPSS 1993 − 2016
Lead time (month)
7
6
5
4
3
2
1
1
2
3
4
5
6
7
8
9
10
11
12
Initial month
CRPSS
1.00
0.75
0.50
0.25
0.00

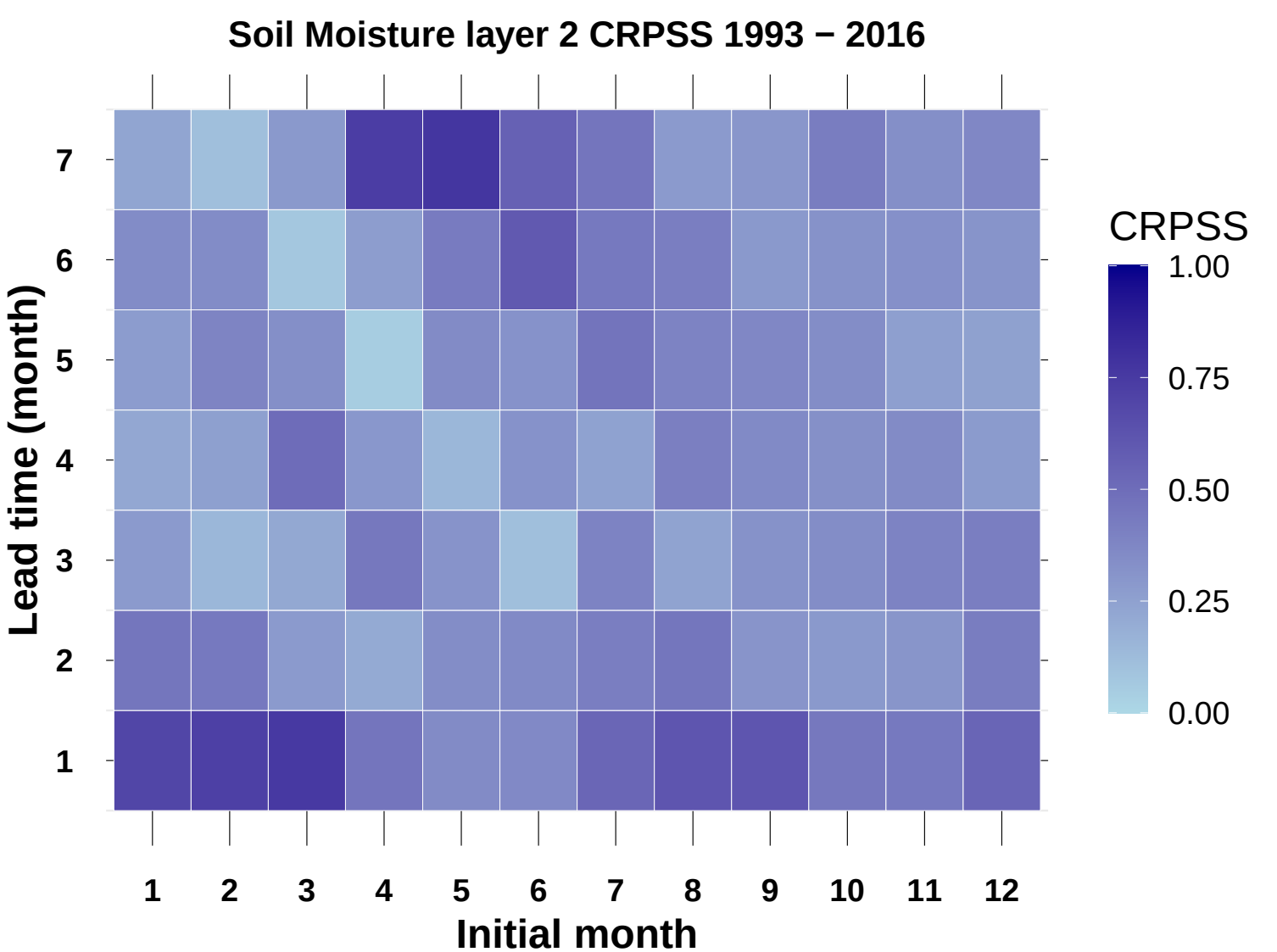

Air temperature (2m) CRPSS 2017 − 2023
Lead time (month)
7
6
5
4
3
2
1
1
2
3
4
5
6
7
8
9
10
11
12
Initial month
CRPSS
1.00
0.75
0.50
0.25
0.00

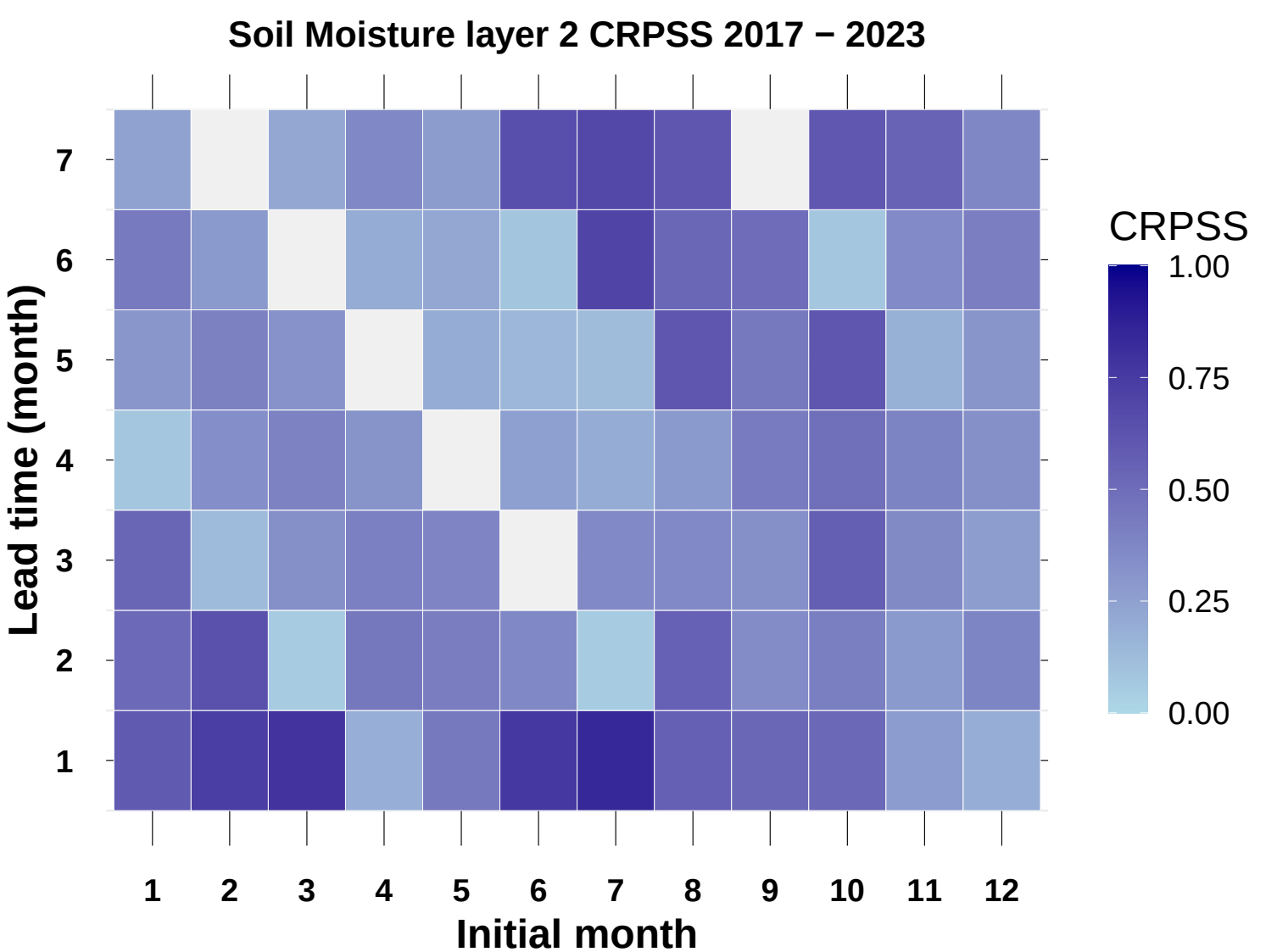

Total precipitation CRPSS 1993 − 2016
Lead time (month)
7
6
5
4
3
2
1
1
2
3
4
5
6
7
8
9
10
11
12
Initial month
CRPSS
1.00
0.75
0.50
0.25
0.00

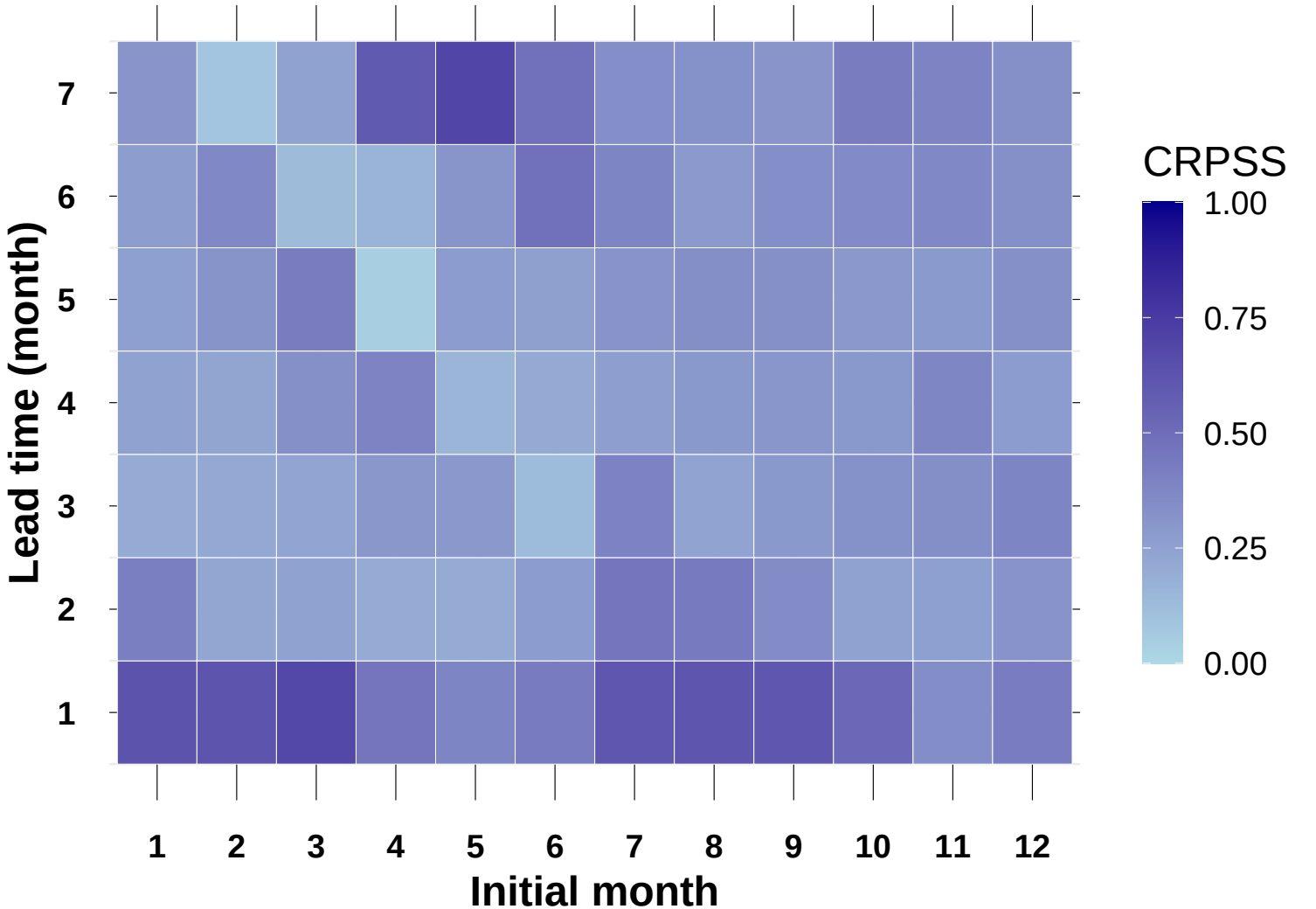

Total precipitation CRPSS 2017 − 2023
Lead time (month)
7
6
5
4
3
2
1
1
2
3
4
5
6
7
8
9
10
11
12
Initial month
CRPSS
1.00
0.75
0.50
0.25
0.00

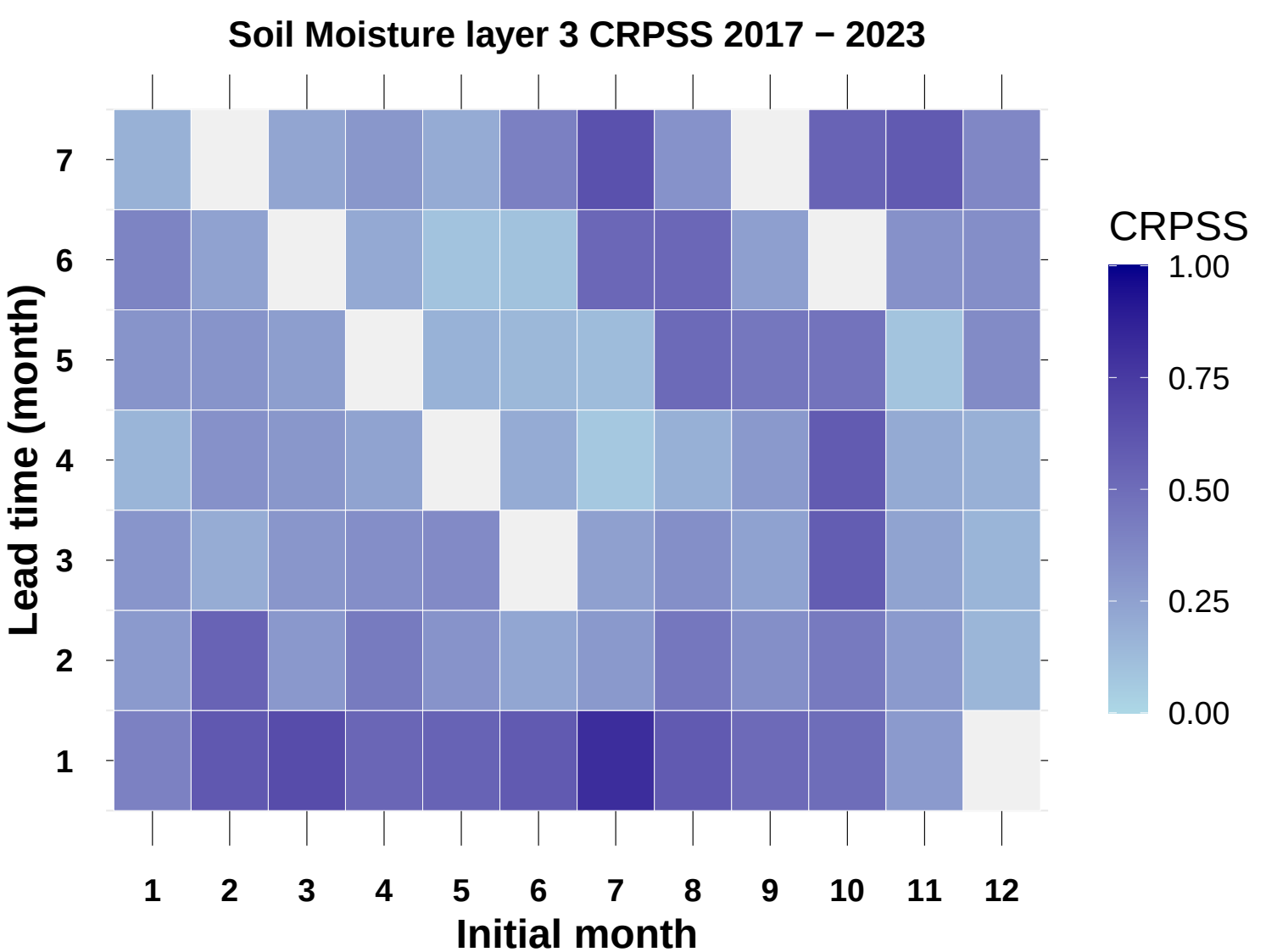

Soil Moisture layer 1 CRPSS 1993 − 2016
Lead time (month)
7
6
5
4
3
2
1
1 2 3 4 5 6 7 8 9 10 11 12
Initial month
CRPSS
1.00
0.75
0.50
0.25
0.00

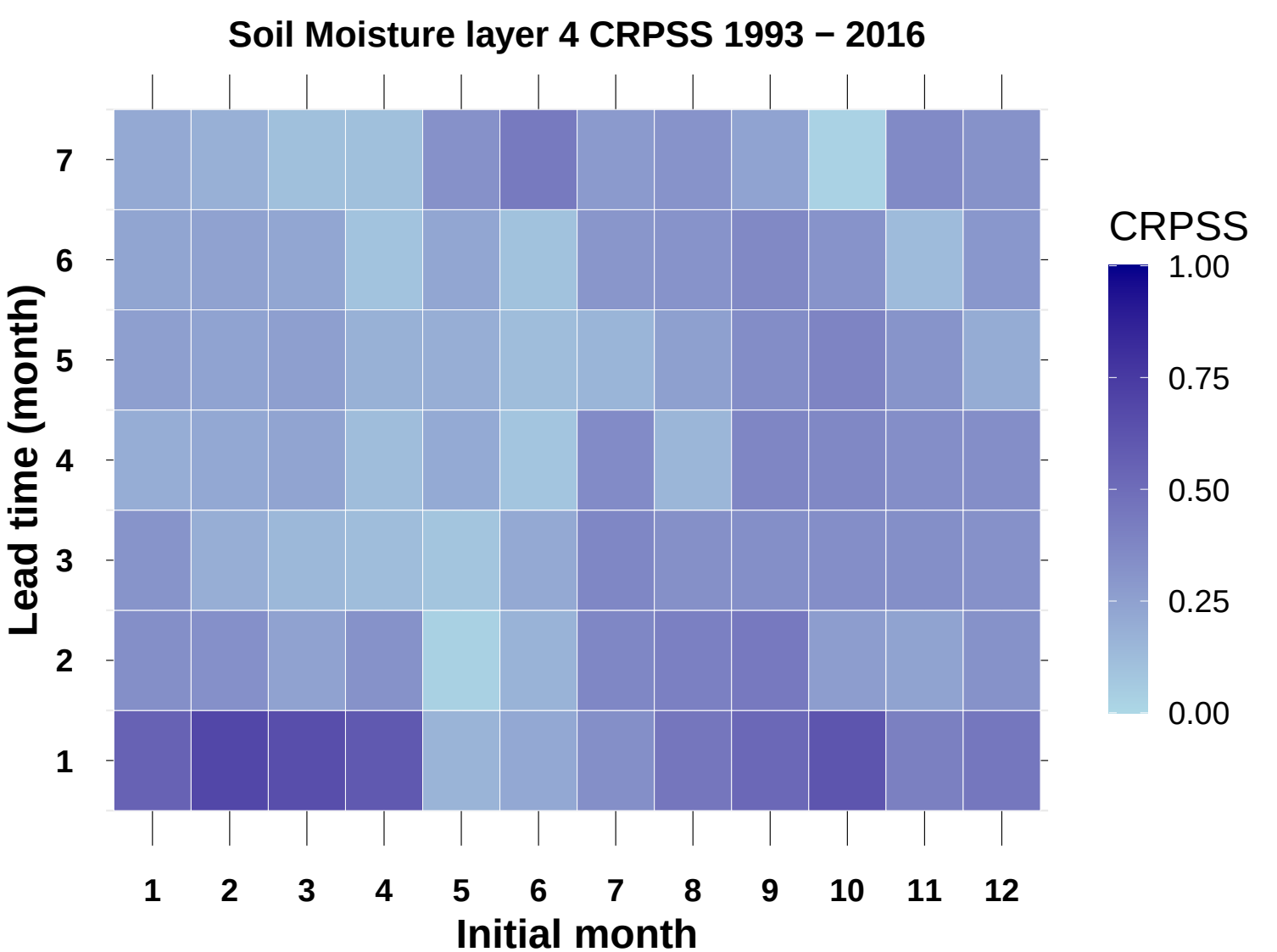

Soil Moisture layer 1 CRPSS 2017 − 2023
Lead time (month)
7
6
5
4
3
2
1
1
2
3
4
5
6
7
8
9
10
11
12
Initial month
CRPSS
1.00
0.75
0.50
0.25
0.00

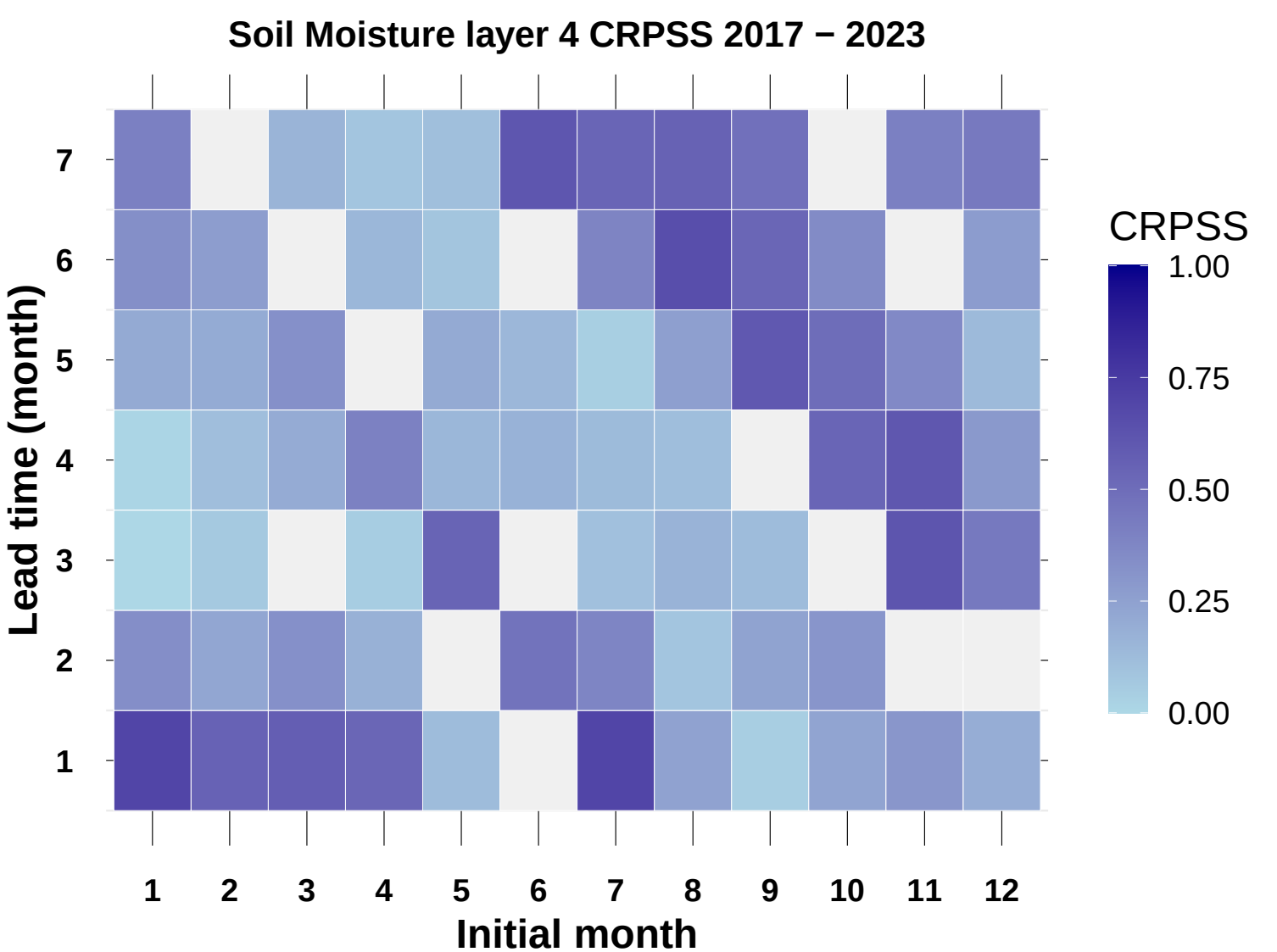

Soil Moisture layer 2 CRPSS 1993 − 2016
Lead time (month)
7
6
5
4
3
2
1
1
2
3
4
5
6
7
8
9
10
11
12
Initial month
CRPSS
1.00
0.75
0.50
0.25
0.00

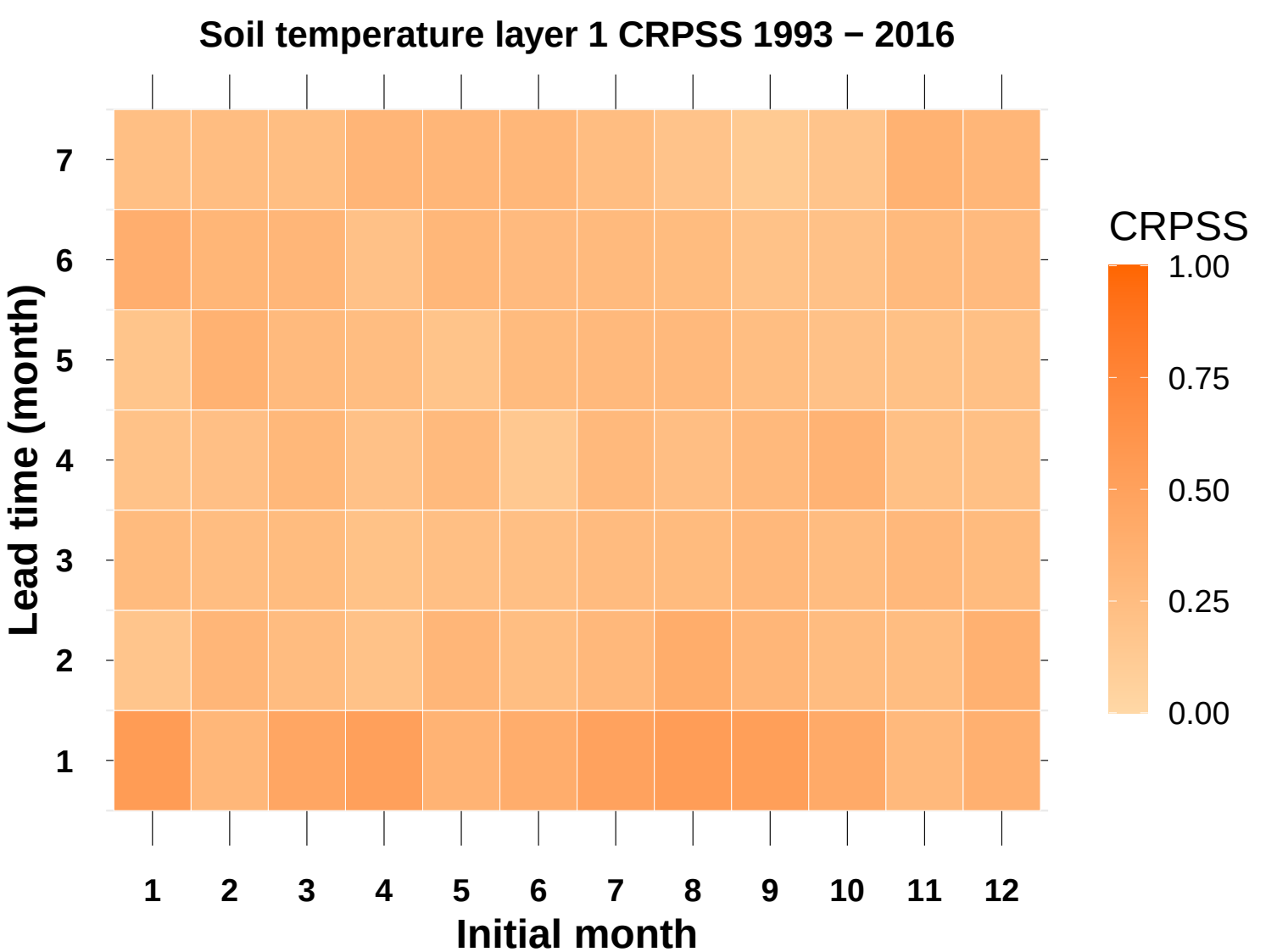

Soil Moisture layer 2 CRPSS 2017 − 2023
Lead time (month)
7
6
5
4
3
2
1
1
2
3
4
5
6
7
8
9
10
11
12
Initial month
CRPSS
1.00
0.75
0.50
0.25
0.00

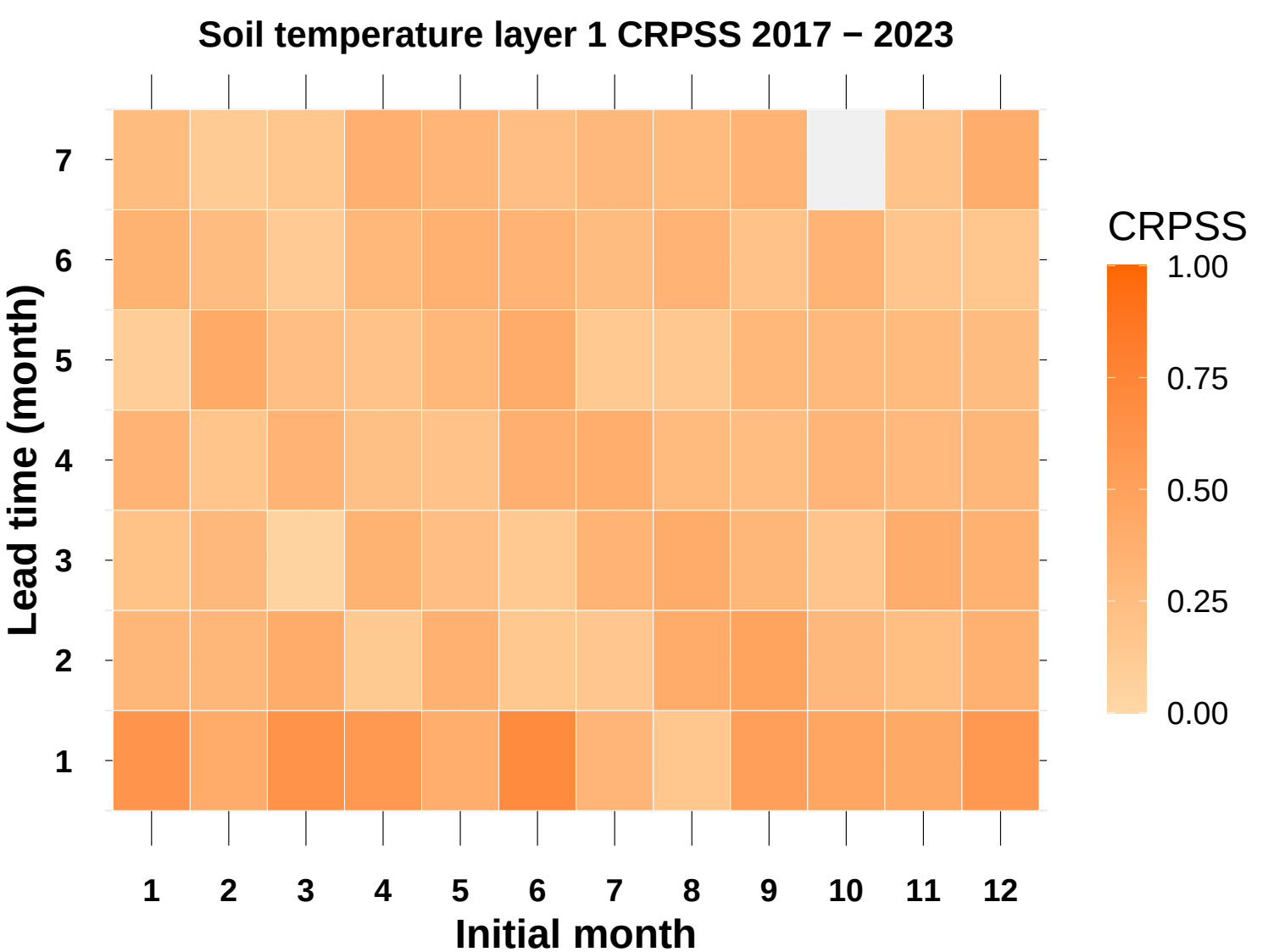

Soil Moisture layer 3 CRPSS 1993 − 2016
Lead time (month)
7
6
5
4
3
2
1
1 2 3 4 5 6 7 8 9 10 11 12
Initial month
CRPSS
1.00
0.75
0.50
0.25
0.00

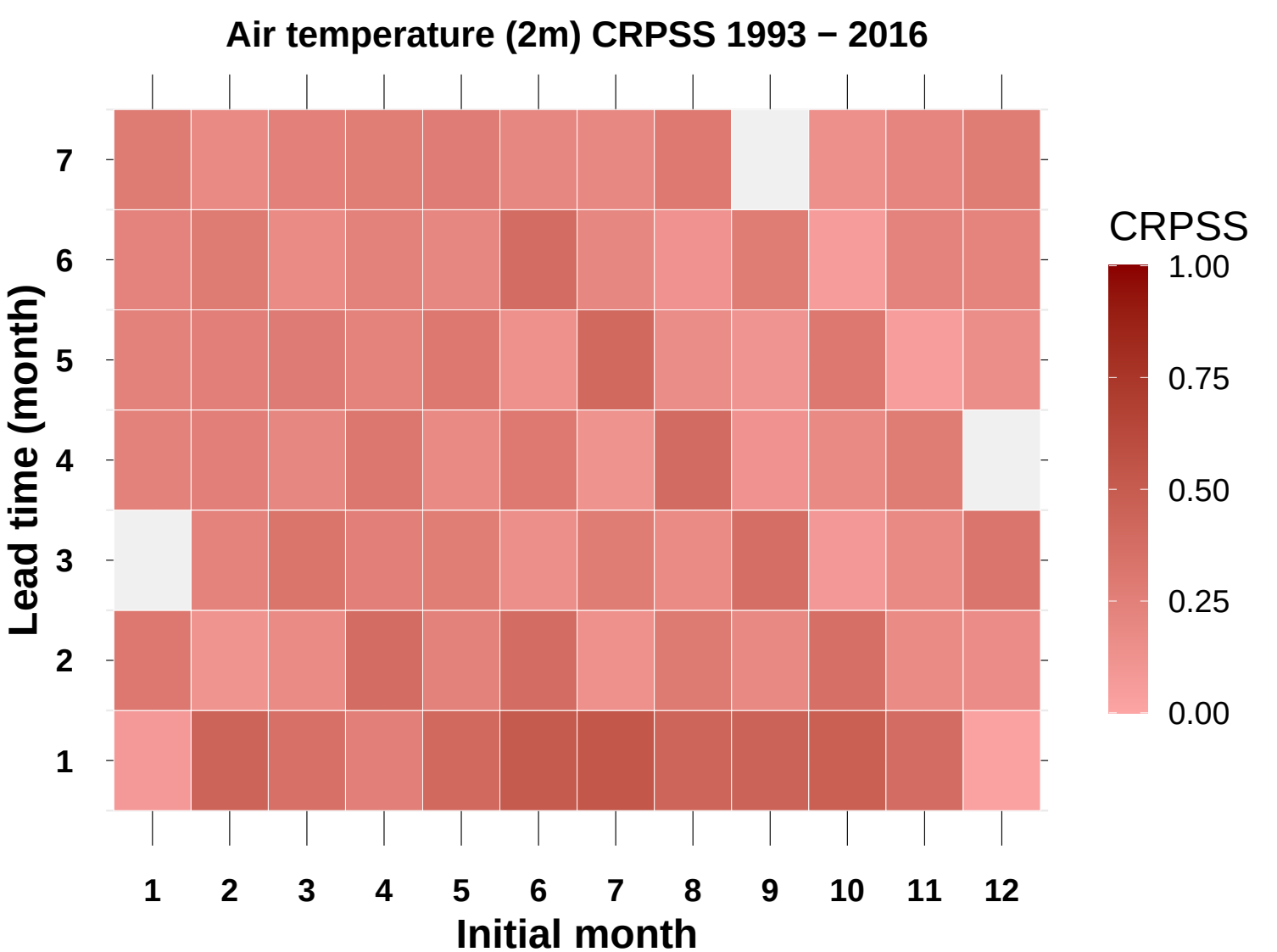

Soil Moisture layer 3 CRPSS 2017 − 2023
Lead time (month)
7
6
5
4
3
2
1
1
2
3
4
5
6
7
8
9
10
11
12
Initial month
CRPSS
1.00
0.75
0.50
0.25
0.00

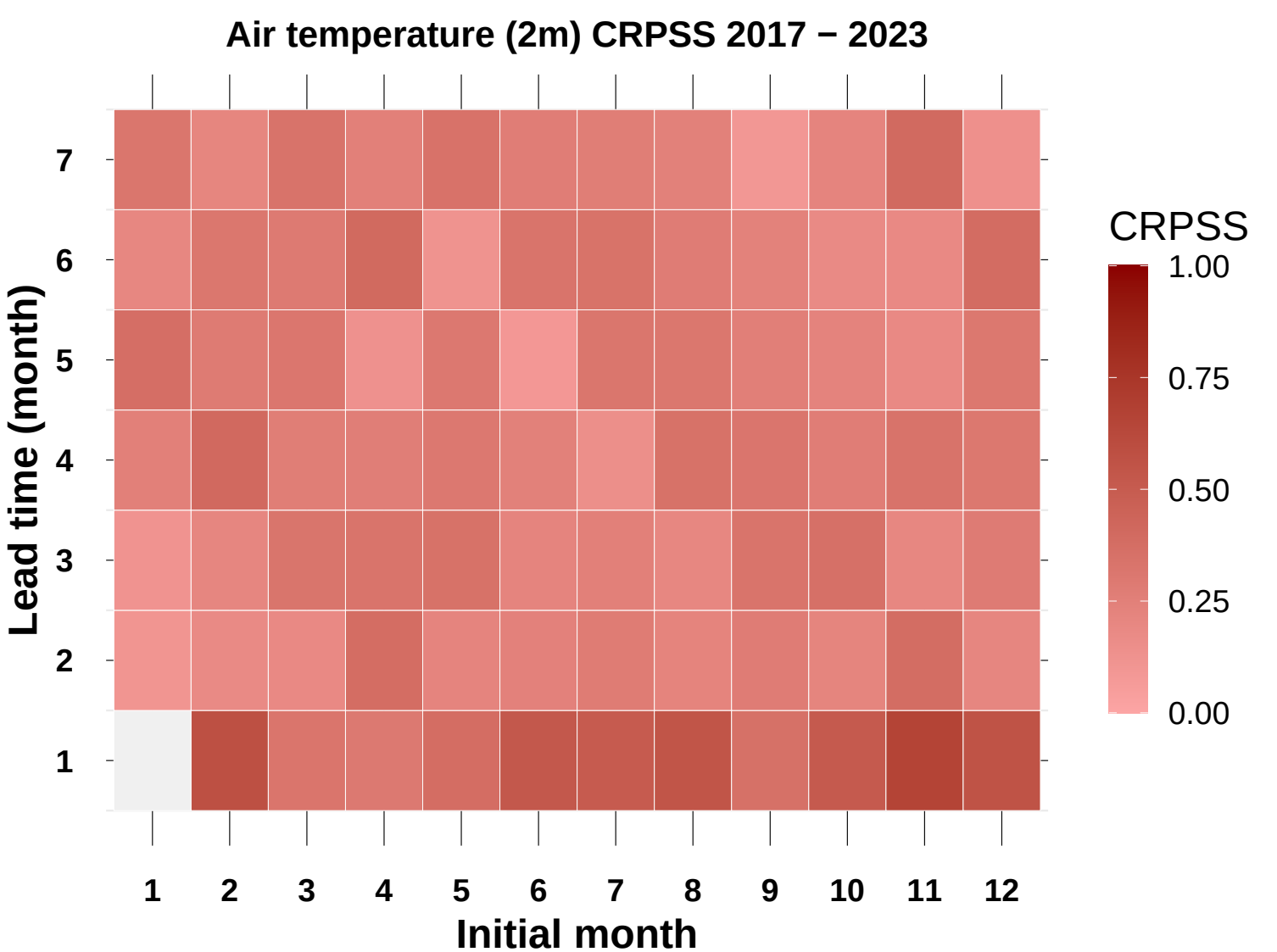

Soil Moisture layer 4 CRPSS 1993 − 2016
Lead time (month)
7
6
5
4
3
2
1
1
2
3
4
5
6
7
8
9
10
11
12
Initial month
CRPSS
1.00
0.75
0.50
0.25
0.00

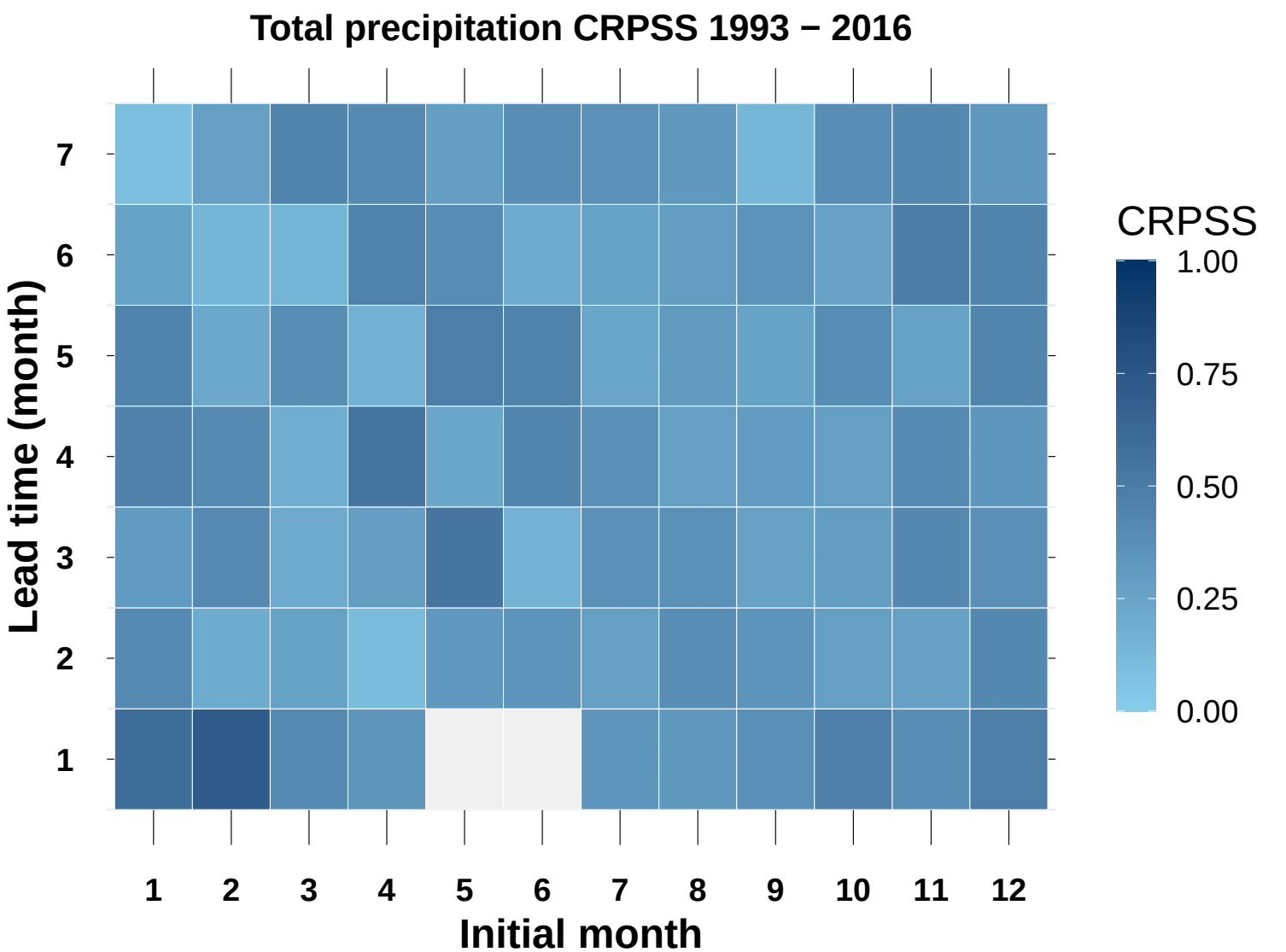

Soil Moisture layer 4 CRPSS 2017 − 2023
Lead time (month)
7
6
5
4
3
2
1
1
2
3
4
5
6
7
8
9
10
11
12
Initial month
CRPSS
1.00
0.75
0.50
0.25
0.00

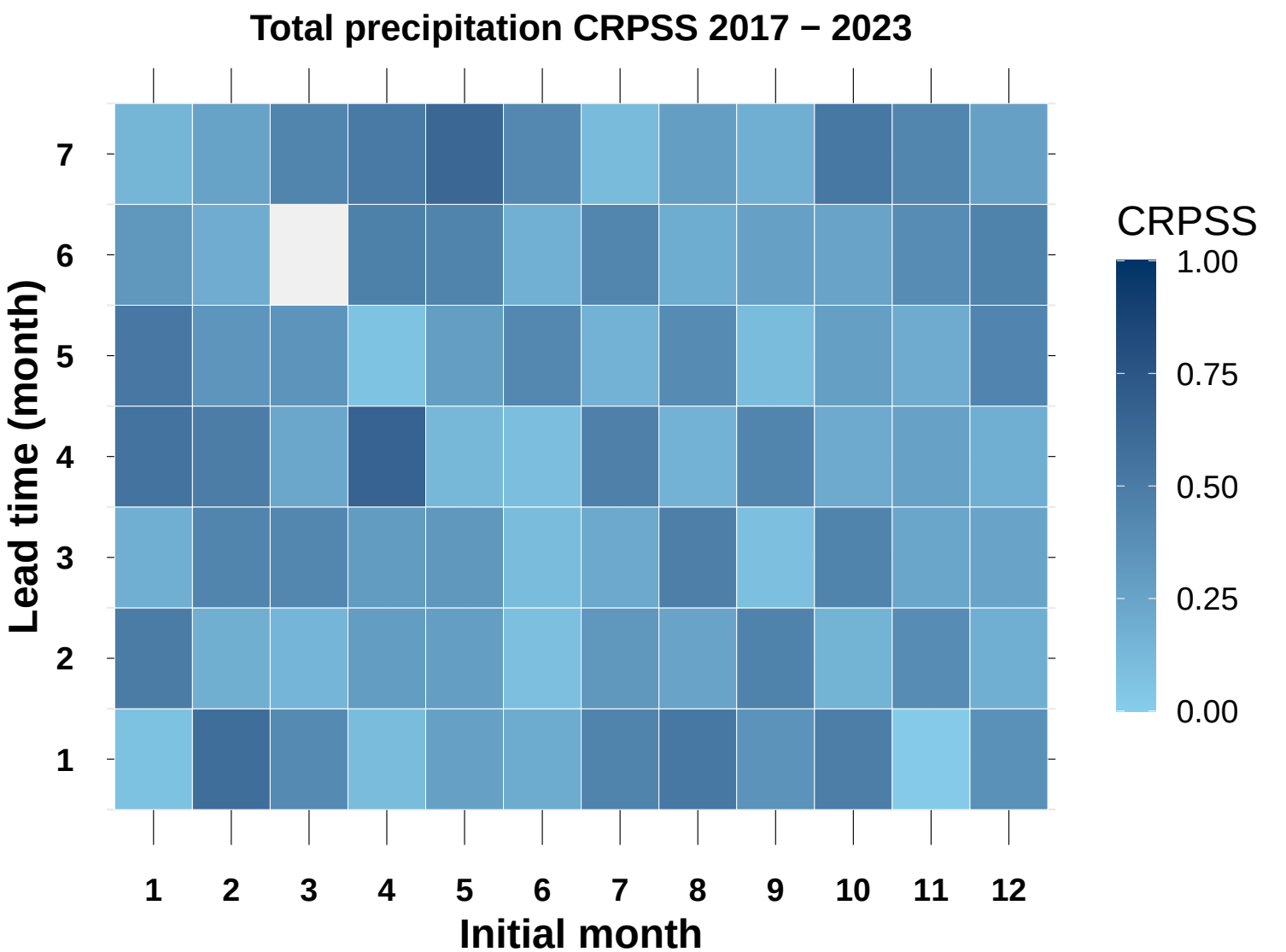

Soil temperature layer 1 CRPSS 1993 − 2016
Lead time (month)
7
6
5
4
3
2
1
1
2
3
4
5
6
7
8
9
10
11
12
Initial month
CRPSS
1.00
0.75
0.50
0.25
0.00

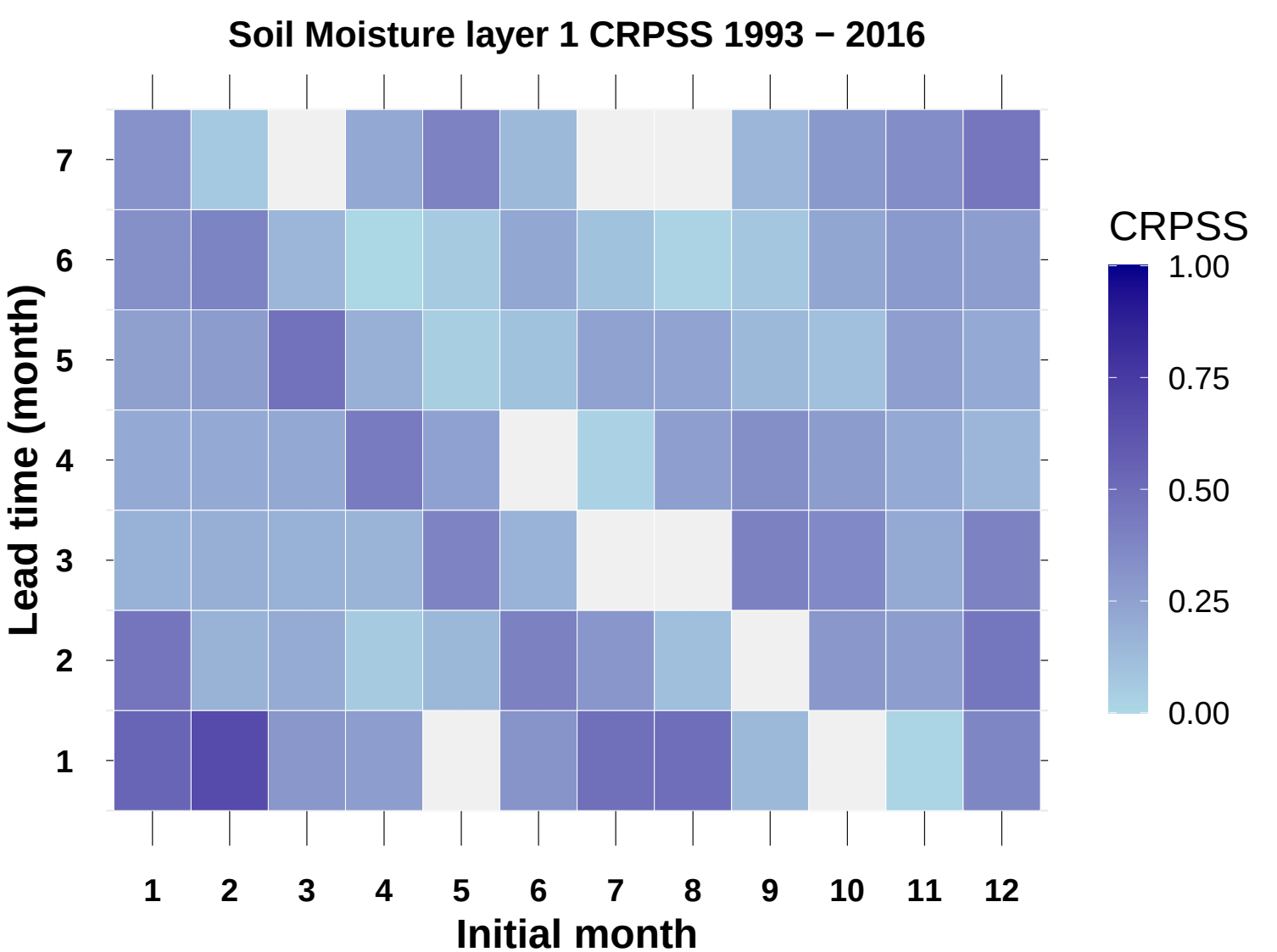

Soil temperature layer 1 CRPSS 2017 − 2023
Lead time (month)
7
6
5
4
3
2
1
1
2
3
4
5
6
7
8
9
10
11
12
Initial month
CRPSS
1.00
0.75
0.50
0.25
0.00

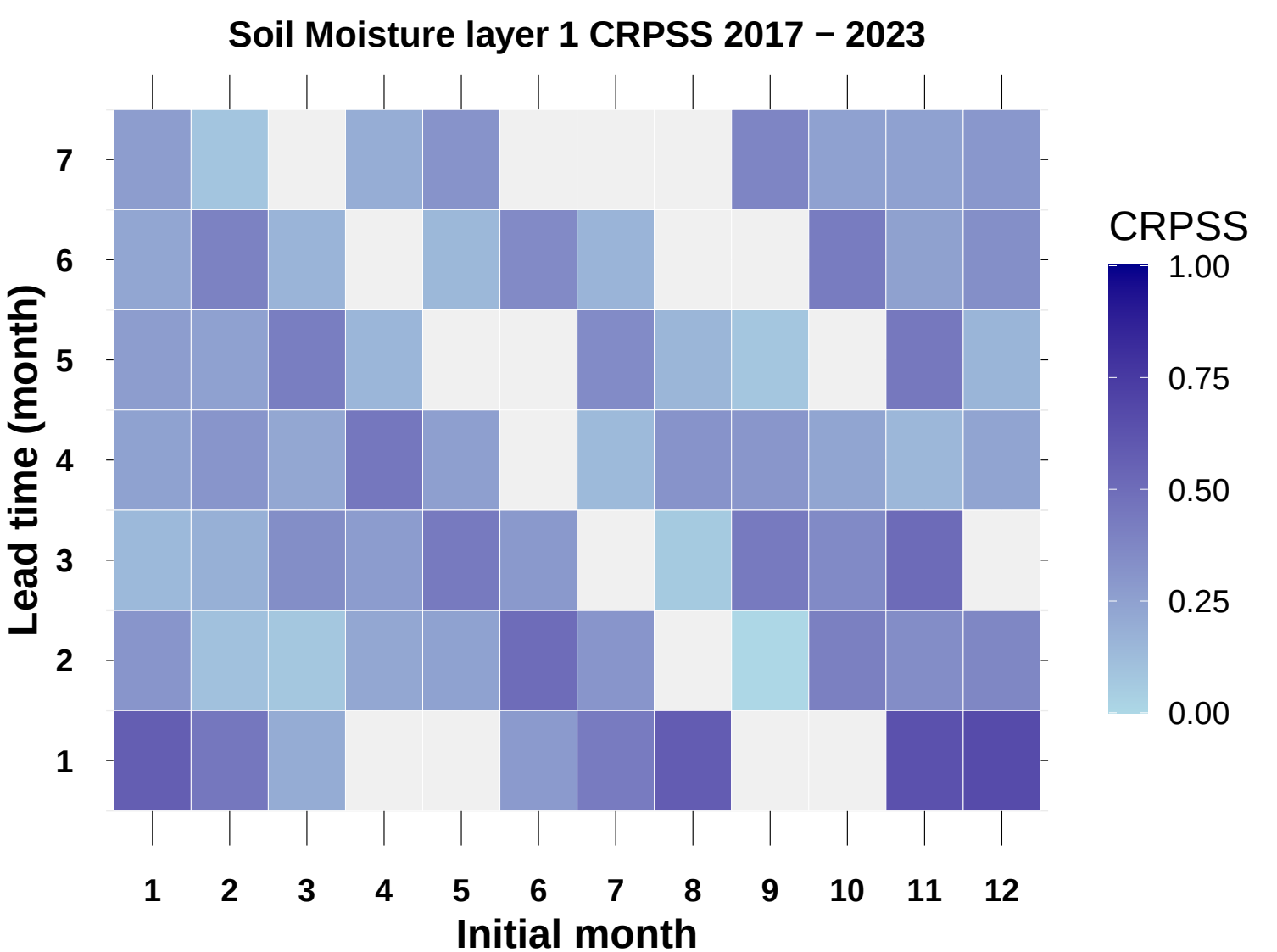

Air temperature (2m) CRPSS 1993 − 2016
Lead time (month)
7
6
5
4
3
2
1
1
2
3
4
5
6
7
8
9
10
11
12
Initial month
CRPSS
1.00
0.75
0.50
0.25
0.00

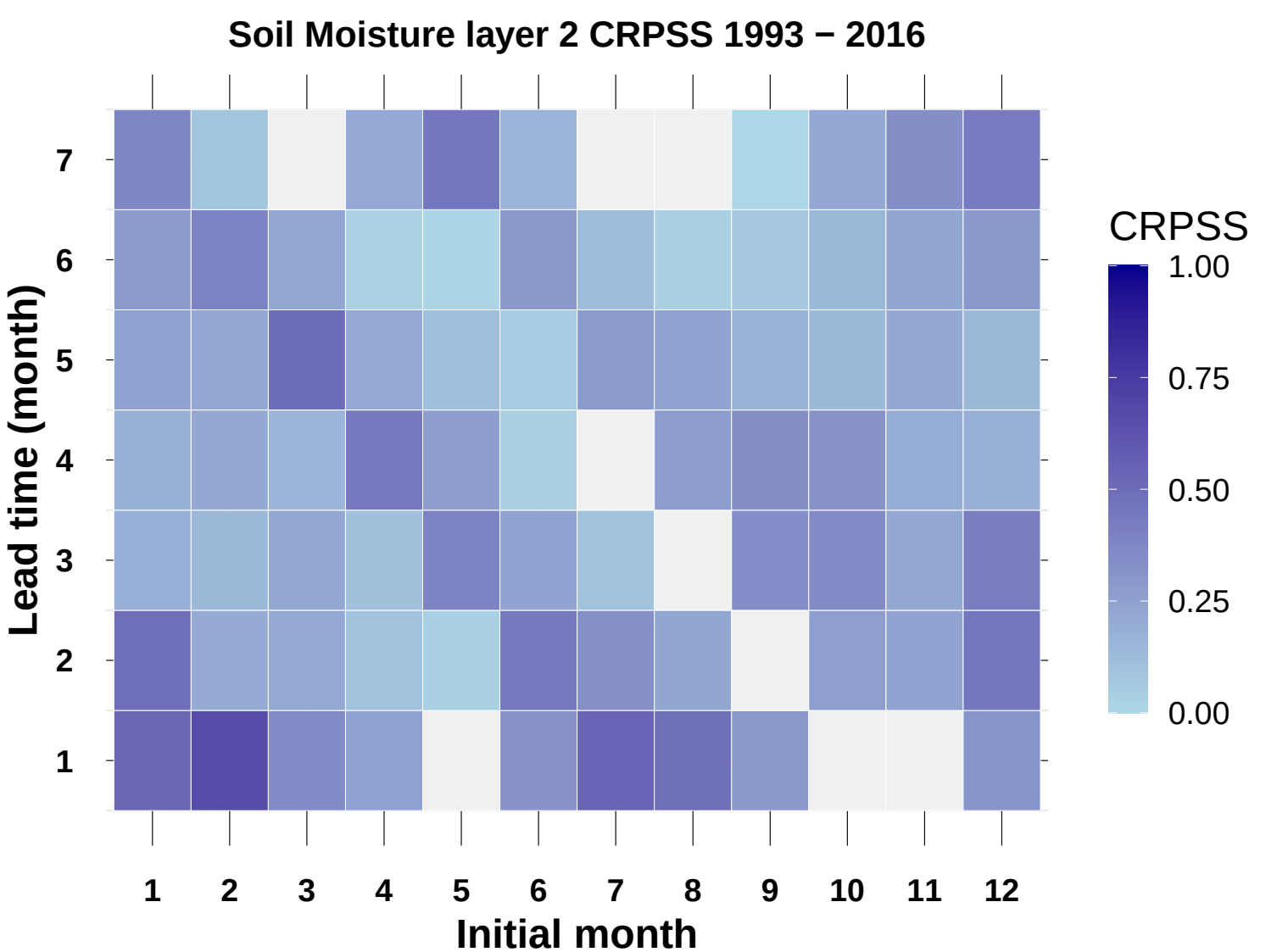

Air temperature (2m) CRPSS 2017 − 2023
Lead time (month)
7
6
5
4
3
2
1
1
2
3
4
5
6
7
8
9
10
11
12
Initial month
CRPSS
1.00
0.75
0.50
0.25
0.00

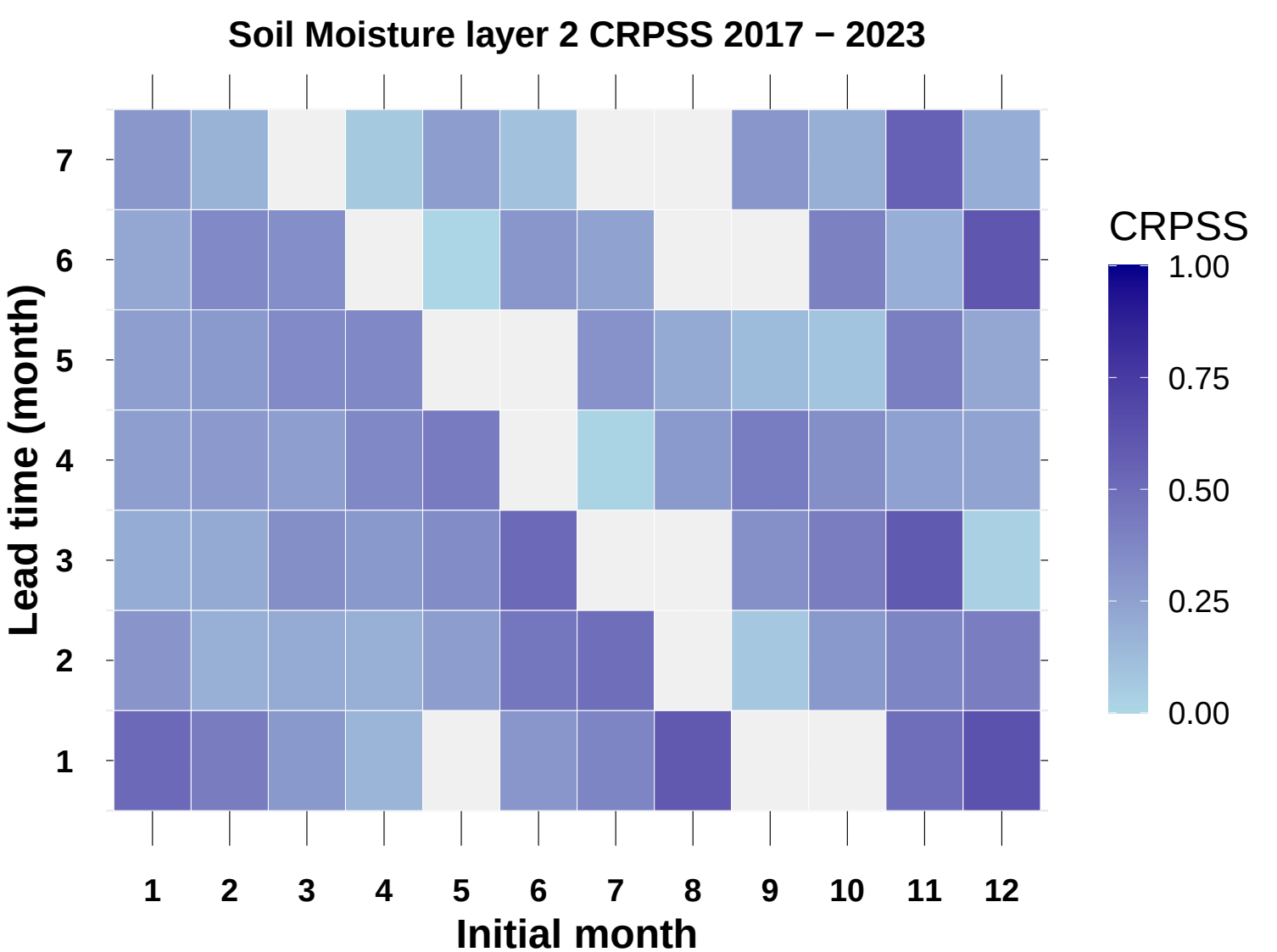

Total precipitation CRPSS 1993 − 2016
Lead time (month)
7
6
5
4
3
2
1
1
2
3
4
5
6
7
8
9
10
11
12
Initial month
CRPSS
1.00
0.75
0.50
0.25
0.00

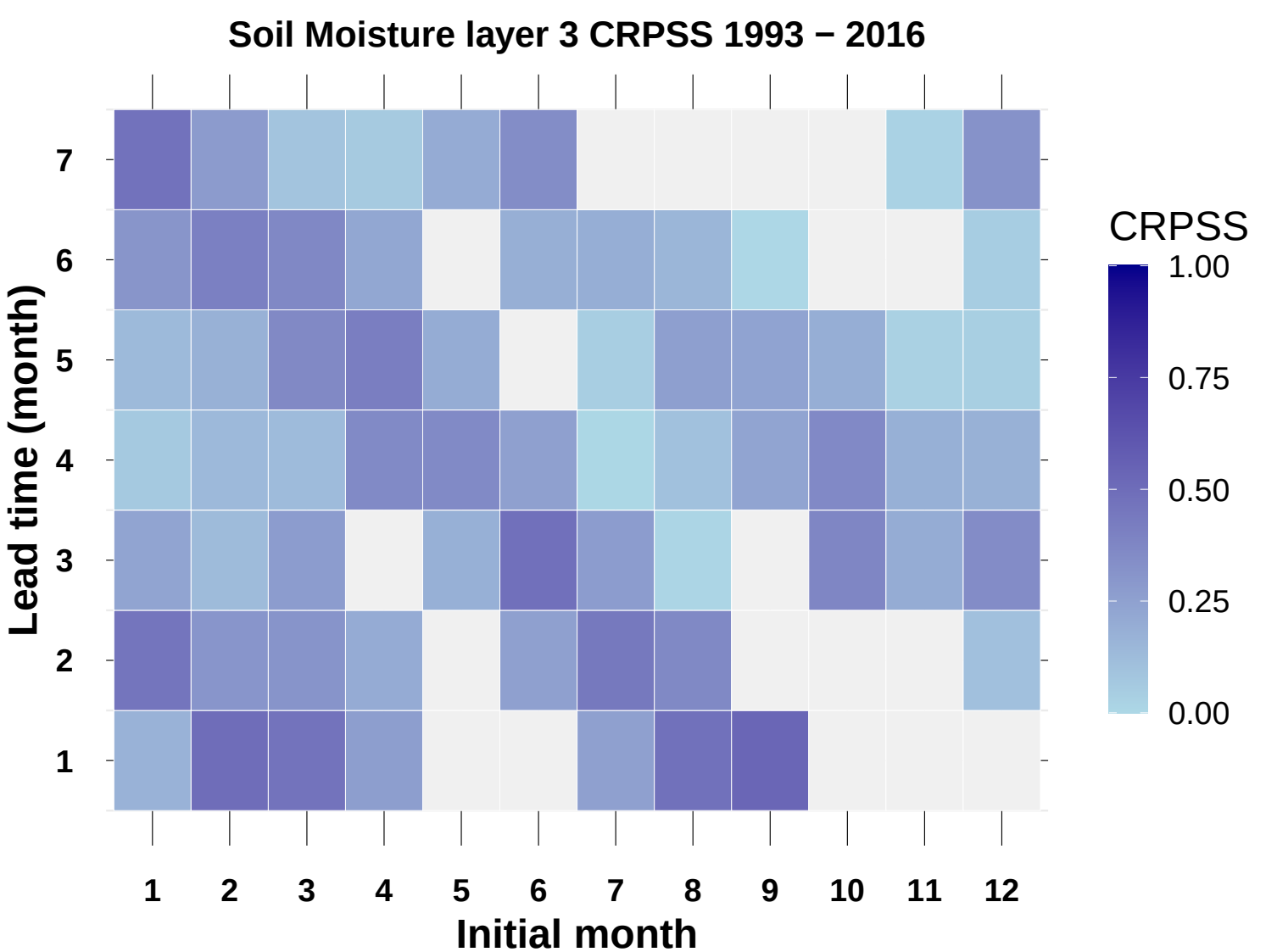

Total precipitation CRPSS 2017 − 2023
Lead time (month)
7
6
5
4
3
2
1
1
2
3
4
5
6
7
8
9
10
11
12
Initial month
CRPSS
1.00
0.75
0.50
0.25
0.00

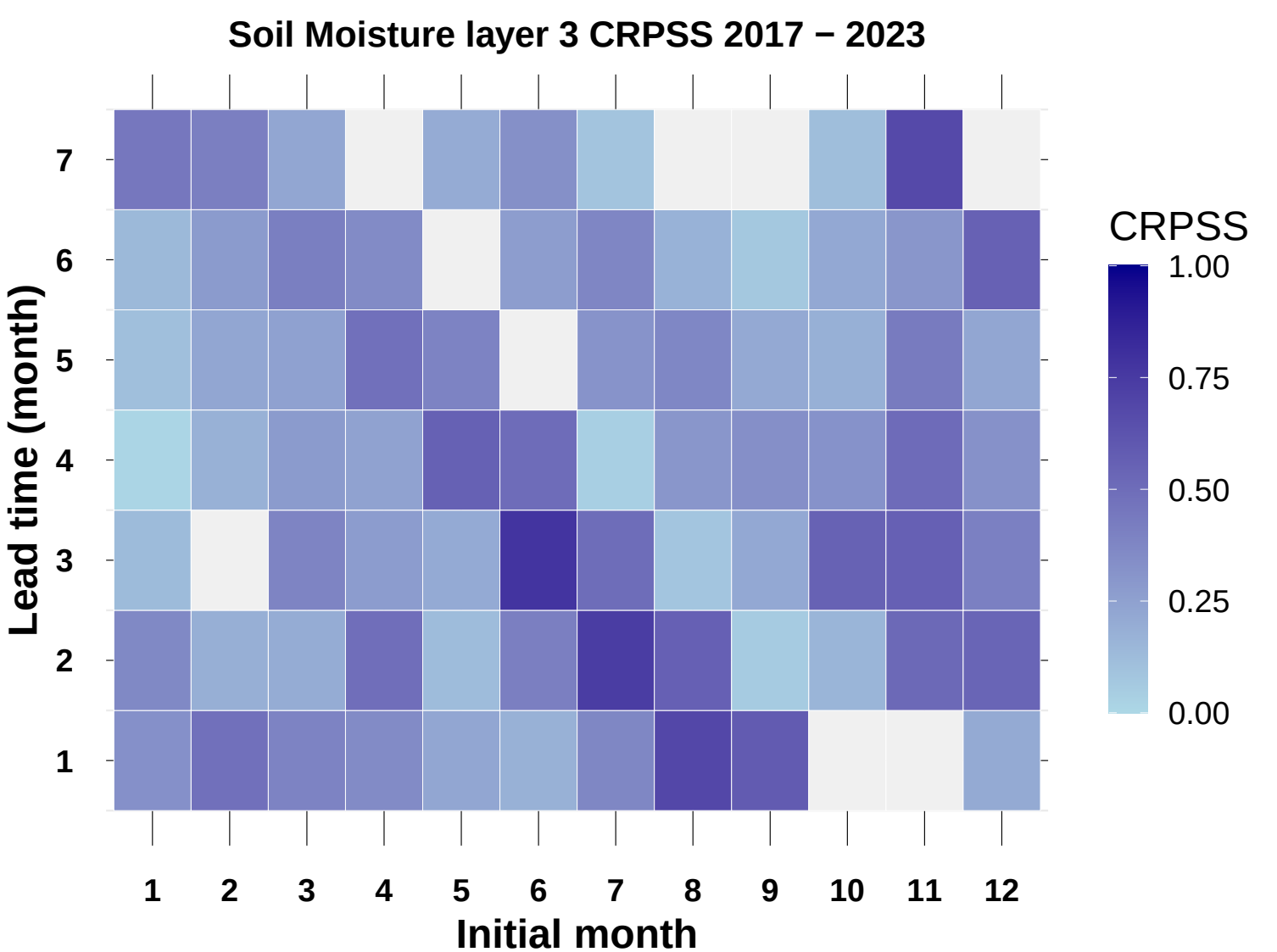

Soil Moisture layer 1 CRPSS 1993 − 2016
Lead time (month)
7
6
5
4
3
2
1
1
2
3
4
5
6
7
8
9
10
11
12
Initial month
CRPSS
1.00
0.75
0.50
0.25
0.00

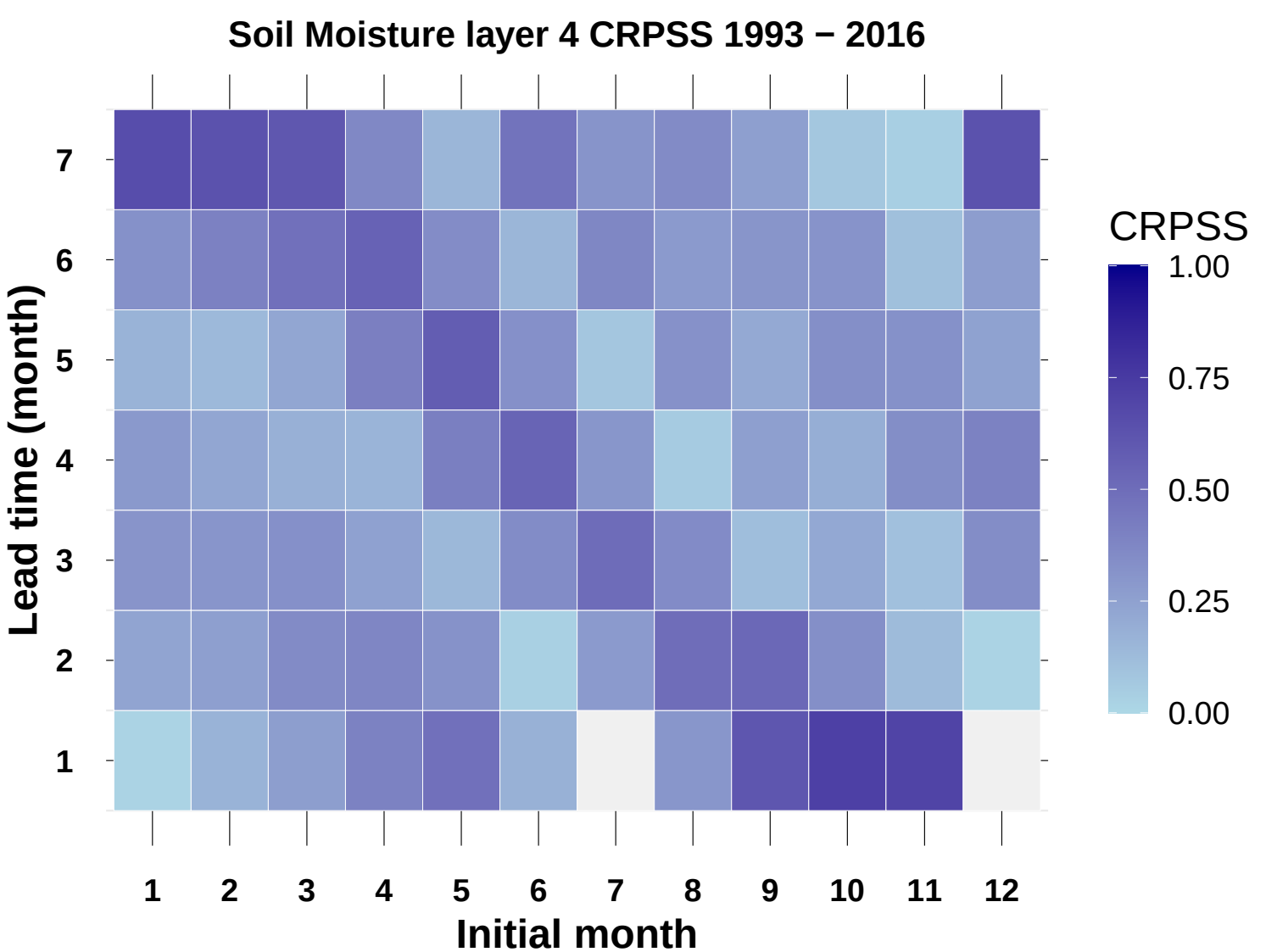

Soil Moisture layer 1 CRPSS 2017 − 2023
Lead time (month)
7
6
5
4
3
2
1
1
2
3
4
5
6
7
8
9
10
11
12
Initial month
CRPSS
1.00
0.75
0.50
0.25
0.00

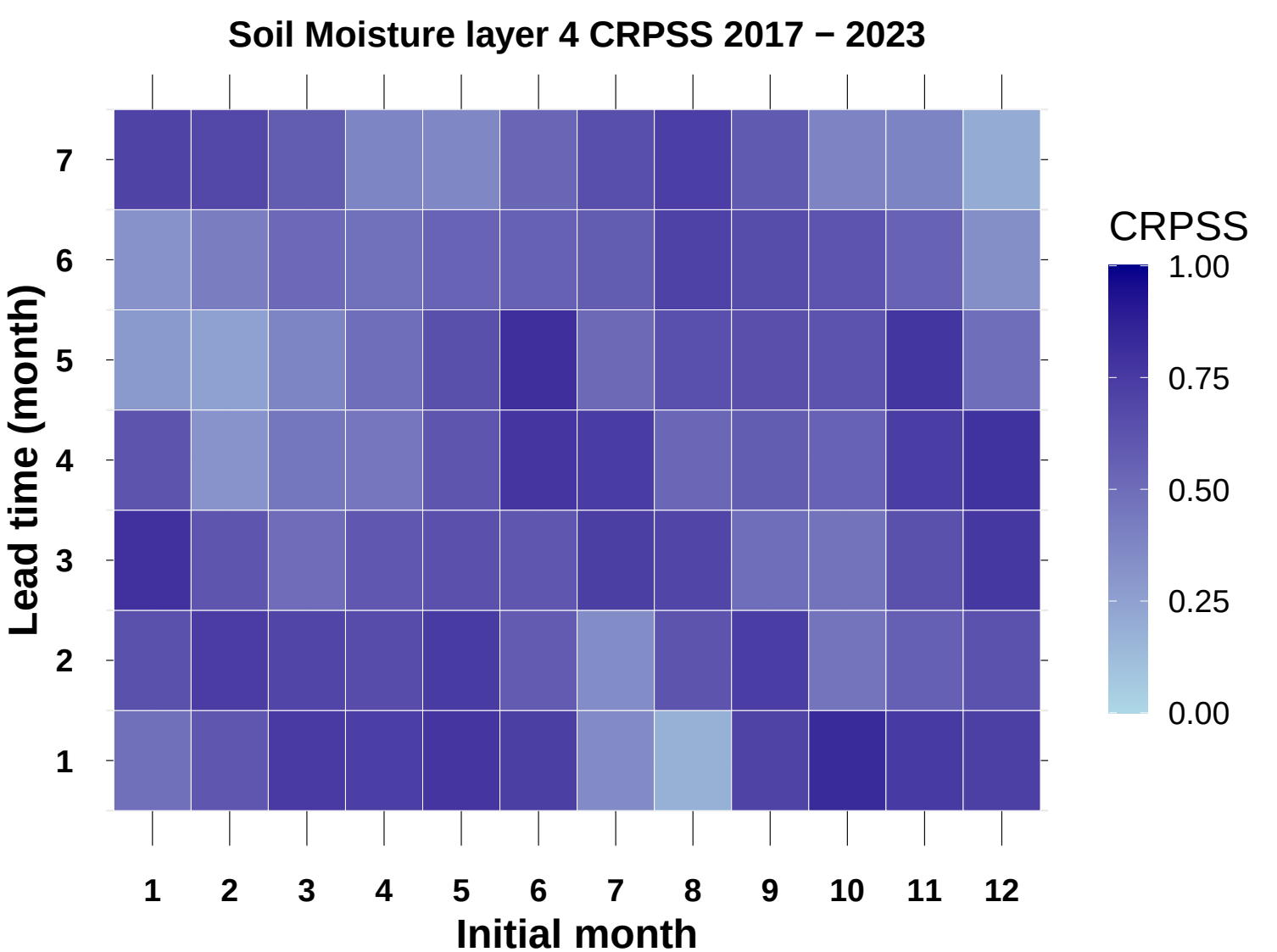

Soil Moisture layer 2 CRPSS 1993 − 2016
Lead time (month)
7
6
5
4
3
2
1
1
2
3
4
5
6
7
8
9
10
11
12
Initial month
CRPSS
1.00
0.75
0.50
0.25
0.00

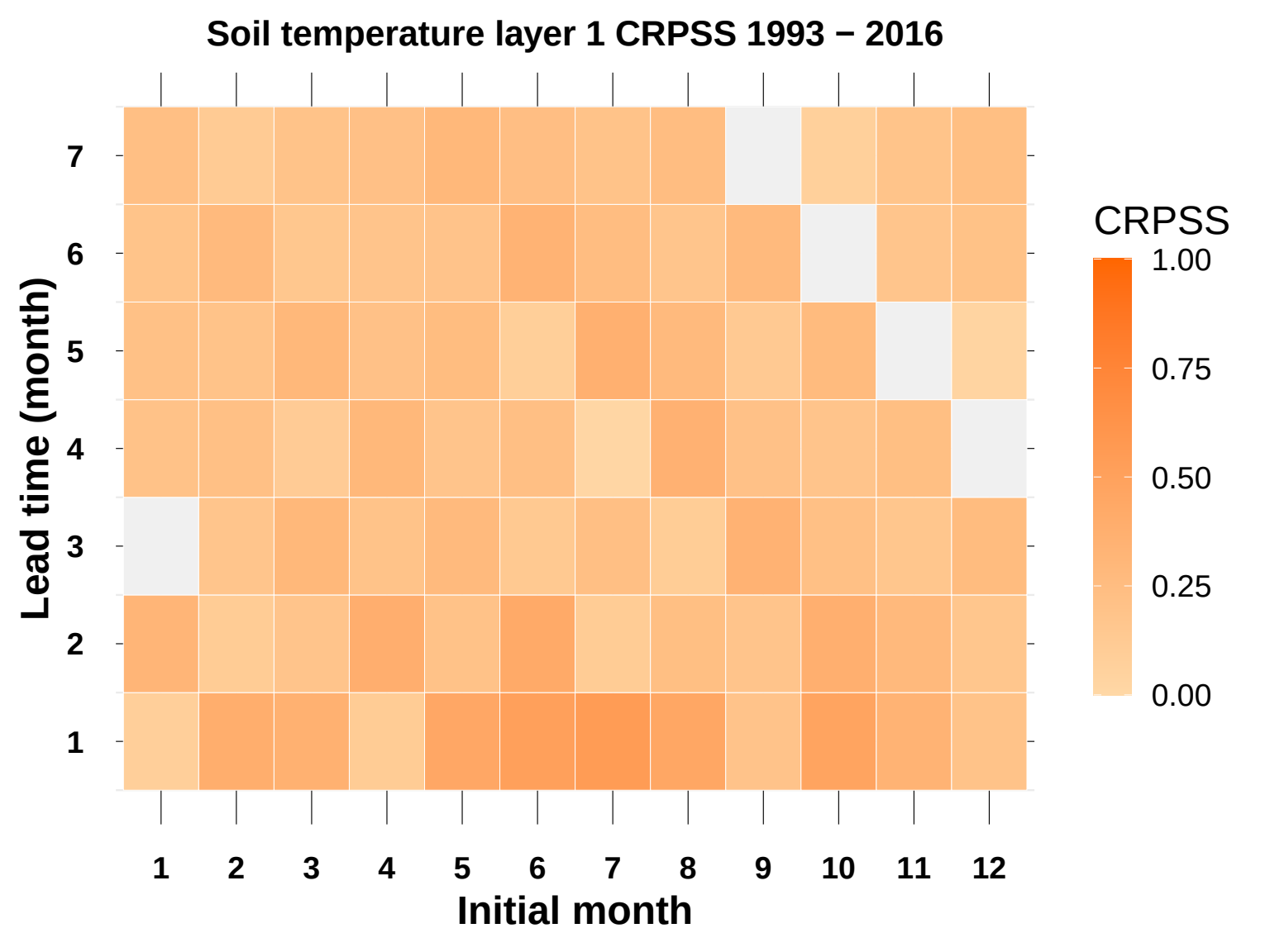

Soil Moisture layer 2 CRPSS 2017 − 2023
Lead time (month)
7
6
5
4
3
2
1
1
2
3
4
5
6
7
8
9
10
11
12
Initial month
CRPSS
1.00
0.75
0.50
0.25
0.00

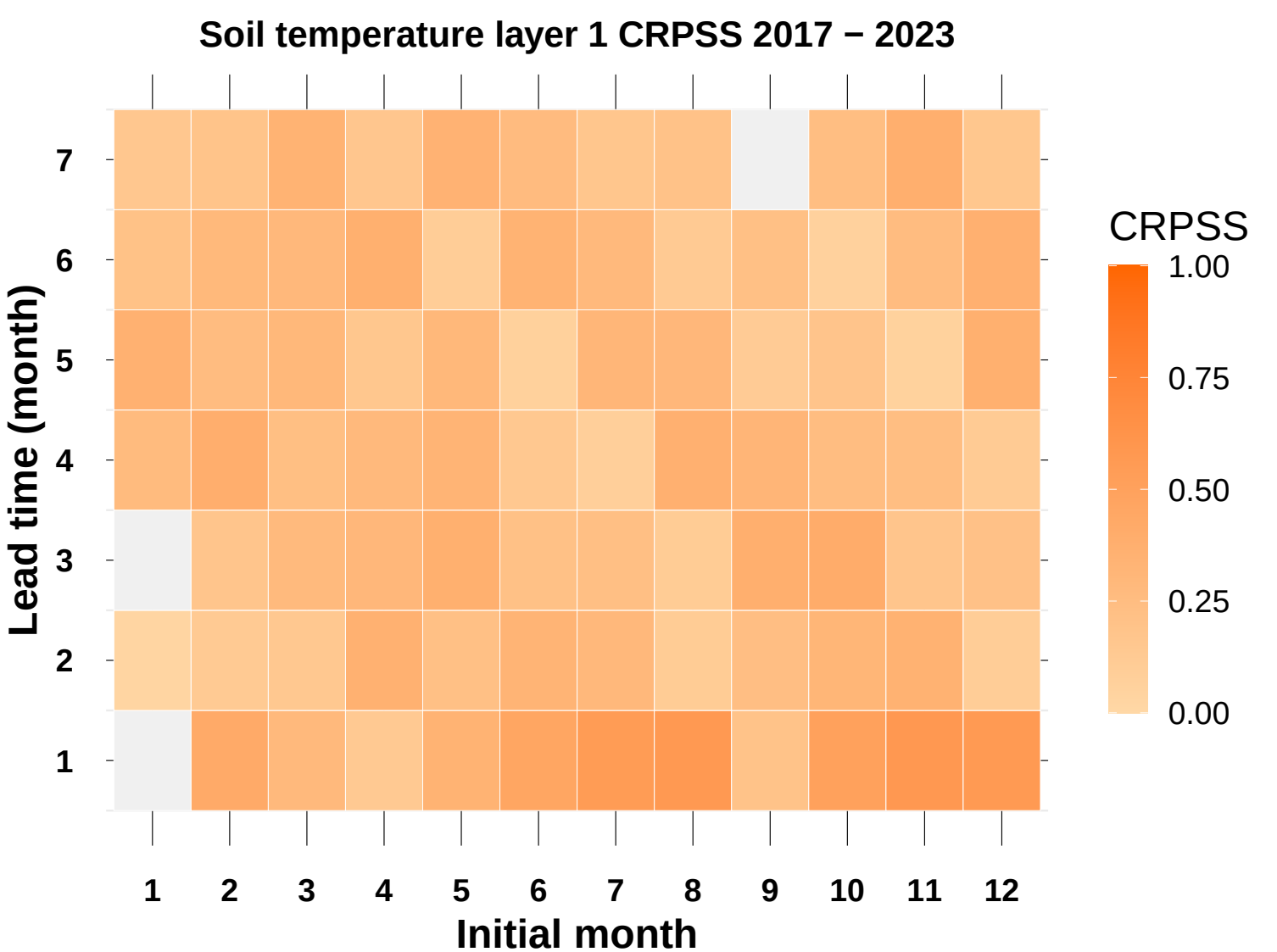

Soil Moisture layer 3 CRPSS 1993 − 2016
Lead time (month)
7
6
5
4
3
2
1
1
2
3
4
5
6
7
8
9
10
11
12
Initial month
CRPSS
1.00
0.75
0.50
0.25
0.00


Soil Moisture layer 3 CRPSS 2017 − 2023
Lead time (month)
7
6
5
4
3
2
1
1
2
3
4
5
6
7
8
9
10
11
12
Initial month
CRPSS
1.00
0.75
0.50
0.25
0.00


Soil Moisture layer 4 CRPSS 1993 − 2016
Lead time (month)
7
6
5
4
3
2
1
1
2
3
4
5
6
7
8
9
10
11
12
Initial month
CRPSS
1.00
0.75
0.50
0.25
0.00


Soil Moisture layer 4 CRPSS 2017 − 2023
Lead time (month)
7
6
5
4
3
2
1
1
2
3
4
5
6
7
8
9
10
11
12
Initial month
CRPSS
1.00
0.75
0.50
0.25
0.00


Soil temperature layer 1 CRPSS 1993 − 2016
Lead time (month)
7
6
5
4
3
2
1
1
2
3
4
5
6
7
8
9
10
11
12
Initial month
CRPSS
1.00
0.75
0.50
0.25
0.00


Soil temperature layer 1 CRPSS 2017 − 2023
Lead time (month)
7
6
5
4
3
2
1
1
2
3
4
5
6
7
8
9
10
11
12
Initial month
CRPSS
1.00
0.75
0.50
0.25
0.00