Response to Referee #1

[Comment 1] The authors of the article by Huang et al aim to improve the simulation of dissolved organic carbon (DOC) in the Yalong River Basin using a coupled modeling framework called SWAT-MDF. The study integrates a process-based hydrological model (SWAT-C) with a data-driven learning module implemented through machine learning algorithms. The authors used SHAP and residual analysis to conduct a comprehensive diagnosis of model components and identify the structural sources of uncertainty in DOC simulation.

The authors find that the Bi-LSTM-based calibration showed the most reliable performance in simulating DOC dynamics with an average NSE of 0.67, which improved the original calibrated SWAT-C results slightly (NSE of 0.51). The SHAP-based global interpretation identified DOC_Simulate, TOT_P, and PRE as the most important predictors of DOC, and a residual analysis revealed that LAI, RH, and DOC_Simulate were the most significant contributors to prediction errors.

[Response 1] We thank the referee for the supportive comments. Please see below our responses to each comment.

General Comment:

[Comment 2] Underlying DOC processes: The article provides very little information on which processes are actually driving the DOC exports from the study area, nor does it provide any information about the trends of these exports, which seem to be relevant (Xu et al., 2024). The authors work with DOC loads, which are mainly driven by the streamflow of the catchment. Why not use the measured DOC concentration data? This would allow a better process analysis, in general and specifically for the model. So in a nutshell: Which processes drive the DOC export of the particular catchment and is this process also the relevant one in the model? Only then one can start thinking about improving the model performance.

[Response 2] We thank the reviewer for this constructive comment. In the revised manuscript, following the reviewer's suggestions, we supplemented the trend analysis of dissolved organic carbon export in Section 3.2, please see the details in our response to comment#5. We completely replaced the target variable from load to concentration, please see the details in our response to comment#29. Furthermore, we supplemented the model driving mechanisms in the discussion Section 4.1, please see the details in our response to comment#23.

[Comment 3] Underlying method: The complex method builds on 72 measurement points in one catchment. I acknowledge the author's trials on training the machine learning methods before applying it on the measured data, but still, this is not a sufficiently rich dataset for a machine learning study (Kratzert et al., 2024).

[Response 3] We thank for referee for these thoughtful comments and suggestions.. We clarified that machine learning inputs are SWAT-C calibrated feature variables and the limited data are used solely for fine tuning. We added four years of runoff data to more accurately calibrate the SWAT-C model:

"DOC observations from 2013-2014 were used as training data, with 20% randomly withheld for validation, while observations from 2019-2020 were used for testing. Because DOC observations are limited, the ML models were not trained directly from observations alone. Instead, the input features were mainly derived from long-term simulated feature parameters generated by the SWAT-C model. Given the limited availability of DOC measurements, a transfer learning strategy was adopted. First, models were pretrained on long-term simulated feature parameters from the SWAT-C model (1972-2020) to establish fundamental relationships between inputs and DOC outputs. Second, the pretrained models were fine-tuned using observed data to better match the true data distribution. In this way, the long-term SWAT-C simulations provide physically consistent training features, while the observed DOC data are primarily used to adjust the model to local conditions. This process mitigated overfitting caused by data scarcity by leveraging generalized patterns learned from simulations.
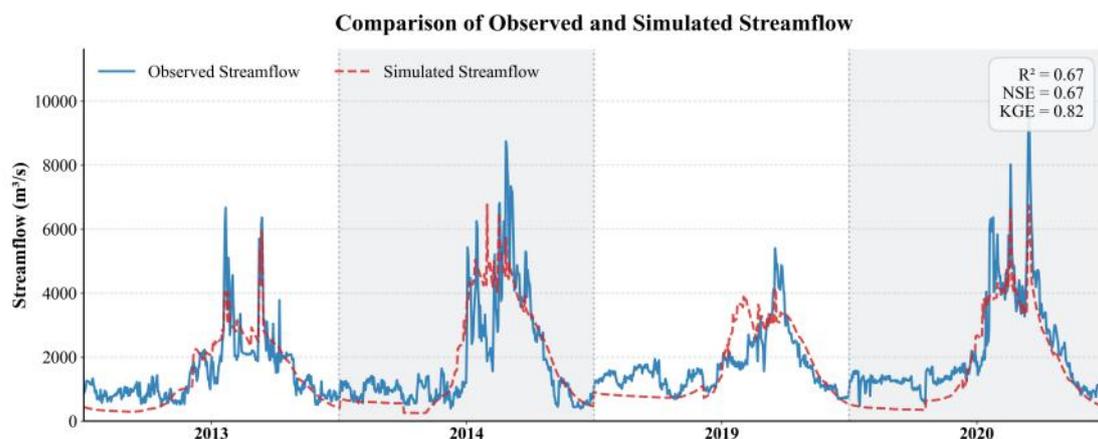
Figure S3. Comparison between observed and simulated daily streamflow used for calibration of the SWAT-C model. "

Specific Comments

[Comment 4] There is a lack of discussion of the limitations of this study. The article mentions some limitations, but it does not provide a comprehensive discussion of the potential limitations and biases of the research.

[Response 4] We thank the reviewer for this comment. We expanded Section 4.4 to comprehensively discuss the study limitations and specific uncertainty sources, as detailed in our response to comment#8.

[Comment 5] The article focuses heavily on the technical details of the methodology and models used, which may be of interest to experts in the field but may not be as relevant to a broader audience. It would rather benefit from studying the processes being actually relevant for the DOC export. Where does it come from, why does it get exported, and is there a trend? If so, can the model reproduce the trend? There are not that many studies out there predicting DOC exports in streams, so why make it overly complicated with the additional layer of machine learning?

[Response5] We thank the referee for this helpful comment. We added Section 3.2 to analyze the spatial distribution and long term temporal trends of dissolved organic carbon export. Furthermore, we clarified the role of the machine learning component as a diagnostic tool grounded in physical processes in the revised manuscript:

"3.2. Spatiotemporal evolution characteristics of simulated DOC

The simulated DOC concentrations in the Yalong River Basin using the SWAT-C model demonstrate significant spatial heterogeneity and pronounced seasonal temporal dynamics(Fig. 4). Spatially, the DOC concentration exhibits a general increasing trend from the northwestern headwaters to the southeastern middle and lower reaches. The northwestern headwater region is characterized by relatively low concentrations, whereas the southeastern downstream region constitutes a high-value zone, with localized maximum concentrations reaching up to 1.75 mg/L. A comparison of the spatial distribution patterns across the years 2000, 2010, and 2020 reveals a gradual contraction in the spatial extent of the high-concentration zones in the lower reaches, accompanied by a trend toward concentration homogenization over time.

Temporally, Figure 4b illustrates the long-term variations in DOC concentrations spanning the period from 2000 to 2020. The monthly concentration curve reveals substantial intra-annual fluctuations, with concentration peaks

occurring consistently during the monsoon season, reaching maximum values exceeding 1.8 mg/L. During the non-monsoon season, the DOC concentration recedes and stabilizes at a baseline level of approximately 0.6 mg/L. The long-term trend line, indicative of interannual variability, demonstrates that the mean annual DOC concentration fluctuates mildly within the range of 0.8 to 1.1 mg/L, notably experiencing a distinct trough period around the year 2012.
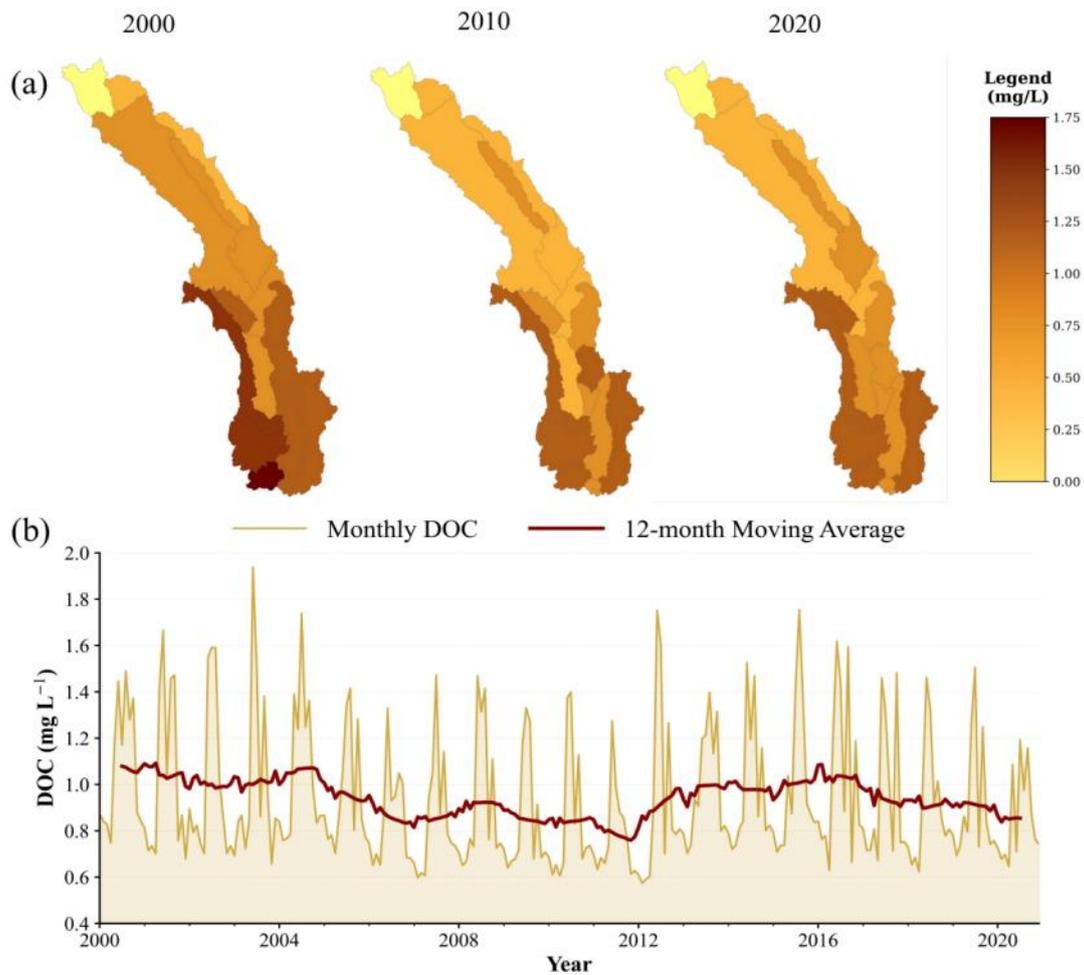
Figure 4. Spatiotemporal evolution of simulated DOC concentrations in the Yalong River Basin. (a) Spatial distribution patterns of DOC concentrations for the representative years 2000, 2010, and 2020. (b) Long-term temporal dynamics of DOC concentrations from 2000 to 2020, illustrating both the monthly fluctuations and the interannual trend."

[Comment 6] Limited discussion of practical implications: The article does not provide a clear discussion of the practical implications of the research findings, such as how they can be applied in real-world settings, how other researchers can benefit from the shown results?

[Response 6] We thank the reviewer for this comment. We expanded Section 4.4 to

discuss the practical value of the SWAT-MDF framework as a diagnostic tool for identifying model structural bottlenecks and serving as a transferable reference for developing future carbon modules in the revised manuscript:

"More importantly, the SWAT-MDF framework offers a systematic diagnostic tool for identifying performance bottlenecks in process-based models. Such diagnostic capability has practical value for model development, as it helps researchers identify which modules or environmental drivers contribute most to simulation errors and therefore where model improvements should be prioritized. Beyond improving DOC simulation in the SWAT-C model, the proposed strategy also shows potential for application in other hydrological and biogeochemical modeling contexts. In principle, the framework is not restricted to SWAT-C and could be extended to models such as SWAT+. Although SWAT+ represents the latest version of the SWAT family, carbon-related functionalities comparable to SWAT-C are still under development. In this context, the SWAT-MDF framework may serve as a transferable diagnostic reference to support the development and validation of future SWAT+ carbon modules and to assist researchers in evaluating and improving carbon simulation modules in other watershed models."

[Comment 7] Lack of transparency in data analysis: The article does not provide a clear description of the data analysis procedures used, which may make it difficult for readers to understand how the results were obtained. There is some information hidden in the supplement material, and the authors also provide their source code (which I highly appreciate), still it remains unclear at key points what has actually been done. There are no results shown of the sensitivity analysis and there is no information on the model calibration (e.g., which algorithm has been used, how many model runs were performed, which performance criterion has been used for the calibration, etc.).

[Response 7] We thank the reviewer for this comment. We clarified the calibration workflow using the SUFI 2 algorithm in SWAT CUP and incorporated parameter sensitivity results, as detailed in our response to comment#28.

[Comment 8] Limited discussion of uncertainty: The article does not provide a clear discussion of the uncertainty associated with the research findings, there are too many places confusions between uncertainty and a lack of model performance. After

reading the article, I have the feeling that the authors interpret uncertainty as some sort of reason for decreasing model performance. However, one should be specific here. There are several sources of uncertainty and they should always be clearly stated. Are we talking about input data uncertainty, model process uncertainty, model parameter uncertainty, model structure uncertainty, etc.?

[Response 8] We thank the reviewer for this comment. We expanded Section 4.4 to explicitly differentiate model performance limitations from specific sources of uncertainty, including input data, model parameters, and model structural uncertainty in the revised manuscript:

"Despite these contributions, several limitations and sources of uncertainty should be acknowledged. First, long term DOC observations in the study basin remain limited, which introduces input data uncertainty and constrains the direct training of data driven models, thereby necessitating the use of simulated variables from the SWAT-C model as input features. Second, parameter uncertainty may arise from the calibration process of the SWAT-C model, as different parameter combinations may yield similar model performance. Third, the current SWAT-C structure simplifies several carbon related processes, such as vegetation derived carbon inputs, soil organic matter decomposition, and their interactions with hydrological transport, which may introduce model structural uncertainty in DOC simulations. Although these uncertainties may influence the interpretation of the results, the combined SHAP and residual analysis still provides a useful diagnostic perspective for identifying key drivers and potential structural weaknesses in the model, offering valuable guidance for future process refinement and model development."

[Comment 9] Lack of clear conclusions and recommendations: The article does not provide clear conclusions and recommendations based on the research findings. What is the main benefit of applying this additional layer of machine learning? What new knowledge can be gained here and in general? There are some elements of this in the manuscript already, but I think a separate section in the discussion and putting the results in the context of available literature would help the readers.

[Response 9] We thank the reviewer for this comment. We expanded Section 4.2 to clarify the specific benefits of the machine learning component and highlighted new insights gained from SHAP and residual analyses in the context of watershed carbon modeling in the revised manuscript:

"Coupling data-driven learning with the process-based SWAT-C model enables these two approaches to complement each other. Within the SWAT-MDF framework, the SWAT-C model provides physically meaningful feature parameters representing watershed processes, which serve as informative inputs for the ML component. The data-driven module then learns nonlinear relationships between these variables and DOC outputs, allowing the framework to capture complex DOC dynamics under varying hydrological and environmental conditions. This hybrid strategy does not merely aim to improve predictive performance; instead, it introduces a data-driven diagnostic layer that helps reveal nonlinear relationships and identify structural bottlenecks within the SWAT-C model."

Response to Technical corrections

[Comment 10] L46: I do not see the argument why the lack of observation data justifies the need for DOC modelling.

[Response 10] We thank the reviewer for this comment. We clarified that modeling serves to provide continuous spatiotemporal estimates of carbon dynamics across the basin to complement limited observational data in the revised manuscript:

"Currently, watershed-scale DOC monitoring predominantly relies on water-quality sampling. However, the limited spatial distribution of fixed monitoring stations and the lack of long-term observational data hinder comprehensive characterization of DOC's spatiotemporal variability (Wang et al., 2025). In this context, process-based modeling approaches can complement field observations by providing spatially and temporally continuous estimates of DOC dynamics, facilitating spatiotemporal analysis of DOC variability and enabling investigation of the underlying hydrological and biogeochemical processes at the basin scale. "

[Comment 11] L60: What is meant by "incomplete module designs and vague parameter representations"? Please provide examples.

[Response 11] We thank the reviewer for this comment. We clarified this statement by providing concrete examples of structural simplifications, such as simplified soil organic matter decomposition and vegetation carbon input representations, in the revised manuscript:

"In addition, structural simplifications in certain model components and

uncertainties in parameter representations may further increase prediction uncertainty. For example, the representation of vegetation-derived carbon inputs, soil organic matter decomposition, and DOC mobilization during hydrological transport is often simplified in process-based watershed models(Qi et al., 2020a). Moreover, several parameters controlling these processes are empirical or poorly constrained by observations, which may introduce additional uncertainty in DOC simulations. "

[Comment 12] L63: Please provide references to this statement.

[Response12] Thank you for pointing this out. In the revised manuscript, we have added an appropriate reference to support this statement.

[Comment 13] L66: Please state again what is meant by the "coupled modelling approach".

[Response 13] We thank the reviewer for this comment. We removed the term and emphasized the need for quantitative diagnostic frameworks to identify key processes and uncertainties in process based watershed models in the revised manuscript:

"Therefore, developing quantitative module diagnosis frameworks is essential for identifying the key processes and sub-modules controlling DOC simulations. Such frameworks can help reveal the sources of model uncertainty and provide guidance for targeted improvements in process-based watershed models."

[Comment14] L68 :Which "variables" are meant here?

[Response 14] We thank the reviewer for this comment. We clarified this sentence by specifying that the variables refer to hydrological, meteorological, and carbon related factors influencing carbon dynamics in the revised manuscript:

"In recent years, machine learning (ML) has demonstrated substantial potential in hydrological modeling, particularly for capturing complex nonlinear interactions among hydrological, meteorological, and carbon-related variables influencing watershed processes (Fan et al., 2020)."

[Comment 15] L75:*Provide reference*

[Response 15] Thank you for pointing this out. In the revised manuscript, we have added an appropriate reference to support this statement.

[Comment 16] L107:Please provide the runoff amount also as an annual average sum in mm to allow for comparison with other catchments.

[Response 16] We thank the reviewer for this comment. We added the annual average runoff depth in mm based on the mean discharge and basin area to facilitate comparison with other catchments in the revised manuscript:.

"The average annual runoff of the basin is 1,914 m³/s, corresponding to an annual runoff depth of about 444 mm yr⁻¹."

[Comment 17] L108:What exactly is meant by particulate matter? How was it measured? Please provide a reference.

[Response 17] We thank the reviewer for this comment. We clarified that particulate matter refers to suspended sediment and added references detailing the measurement methods and data sources from the hydrological station in the revised manuscript:

"the average annual export of suspended sediment matter reaches approximately $2.55*10^{10}$ kg, based on long-term observations from hydrological stations in the basin(Liu et al., 2019)."

[Comment 18] L110:Please provide an overview of the temperature in the study area as well, to give the reader an impression of the local climate.

[Response 18] Following the referee's suggestion, we added a description of the temperature conditions to provide a clearer overview of the regional climate in the revised manuscript: *"The mean annual temperature in the basin ranges from −4.9°C to 19.7°C, decreasing from south to north and with increasing elevation. "*

[Comment 19] L127:State the temporal resolution of the measured DOC data.

[Response 19] Following the referee's suggestion, we clarified the temporal resolution and nature of the DOC observations as discrete measurements collected on specific sampling dates in the revised manuscript:*"DOC observations for 2013–2014 and 2019–2020 were derived from discrete DOC measurements collected on specific sampling dates at the Tongzilin Hydrological Station, as reported by Xu et al., (2024)."*

[Comment 20] L129:Sentence out of context.

[Response 20] Following the referee's suggestion, we removed the disconnected sentence to improve the clarity and coherence of the text in the revised manuscript.

[Comment21] Table 1:Please state the year for the sources of Runoff and Sediment data. Please also try to provide a link where information about this source can be found.

[Response 21] Following the referee's suggestion, we added the temporal coverage and specified the data source as the Hydrological Yearbook of the People's Republic of China, as detailed in our response to comment#22.

[Comment 22] Table 1:The reference of Xu et al 2024 for DOC data does not provide any daily observations, as the table implies here. Xu et al measured monthly, and increased the sampling to weekly in the monsoon season. Additionally, I have my doubts that there are daily sediment observations, which can not be double checked however, as there is no reference.

[Response 22] Following the referee's suggestion, we clarified that DOC data consist of discrete daily measurements on specific sampling dates and specified the official source for daily sediment observations as the Hydrological Yearbook of the People's Republic of China in the revised manuscript:

Data Type	Description	Source
Runoff Data	Discrete daily observations (2013-2014, 2019-2020)	Hydrological Yearbook of the People's Republic of China (http://www.mwr.gov.cn/)
Sediment Data	Discrete daily observations (2013-2014, 2019-2020)	Hydrological Yearbook of the People's Republic of China (http://www.mwr.gov.cn/)
DOC Data	Discrete daily observations (2013-2014, 2019-2020)	Escalating Carbon Export from High-Elevation Rivers in a Warming Climate (Xu et al., 2024) Supplementary data"

[Comment 23] Table 2:There are no parameters shown here, even though the table title suggests it. I think those are model in and outputs? And Table S4 is incomplete

as it shows "etc." at some points, e.g. at the core of the study, the DOC simulation. As such it is not possible to judge whether the selection of the parameters is meaningful. I have my doubts, e.g. because why is the meteorological forcing part of the calibration? I do not see any justification for this in the manuscript.

[Response 23] We thank the referee for this helpful comment, we clarified Table 2 lists simulated state variables as machine learning features rather than calibration parameters, expanded Table S4 with precise source code locations, and justified this variable selection by linking them to physical processes within the newly added Section 4.1 in the revised manuscript:

*"**Table 2.** SWAT-C simulated variables used as machine learning feature variables*

Data Name	Definition	SWAT-C Module
FLOW_OUT	*Streamflow output*	*Runoff module*
SED_OUT	*Sediment output*	*Sediment module*
DOC_Simulate	*Simulated total DOC*	*Carbon cycle module*
BIOM	*Biomass*	*Biomass module*
ET	*Evapotranspiration*	*Evapotranspiration module*
TOT_P	*Total phosphorus output*	*Pollutant transport module*
LAI	*Leaf area index*	*Vegetation growth module*
SW	*Soil water content*	*Soil moisture module*
PCP. RH, TMAX, TMIN, SR, WIND	*Precipitation, Relative humidity, Max/Min Temperature, Solar radiation, Wind speed*	*Meteorological forcing module*

Note: The specific locations of the SWAT-C modules in the source code are provided in Table S4 in the Supplementary Materials. All feature parameters were derived from the calibrated SWAT-C model.

4.1. Spatiotemporal variation of dissolved organic carbon and its driving mechanisms

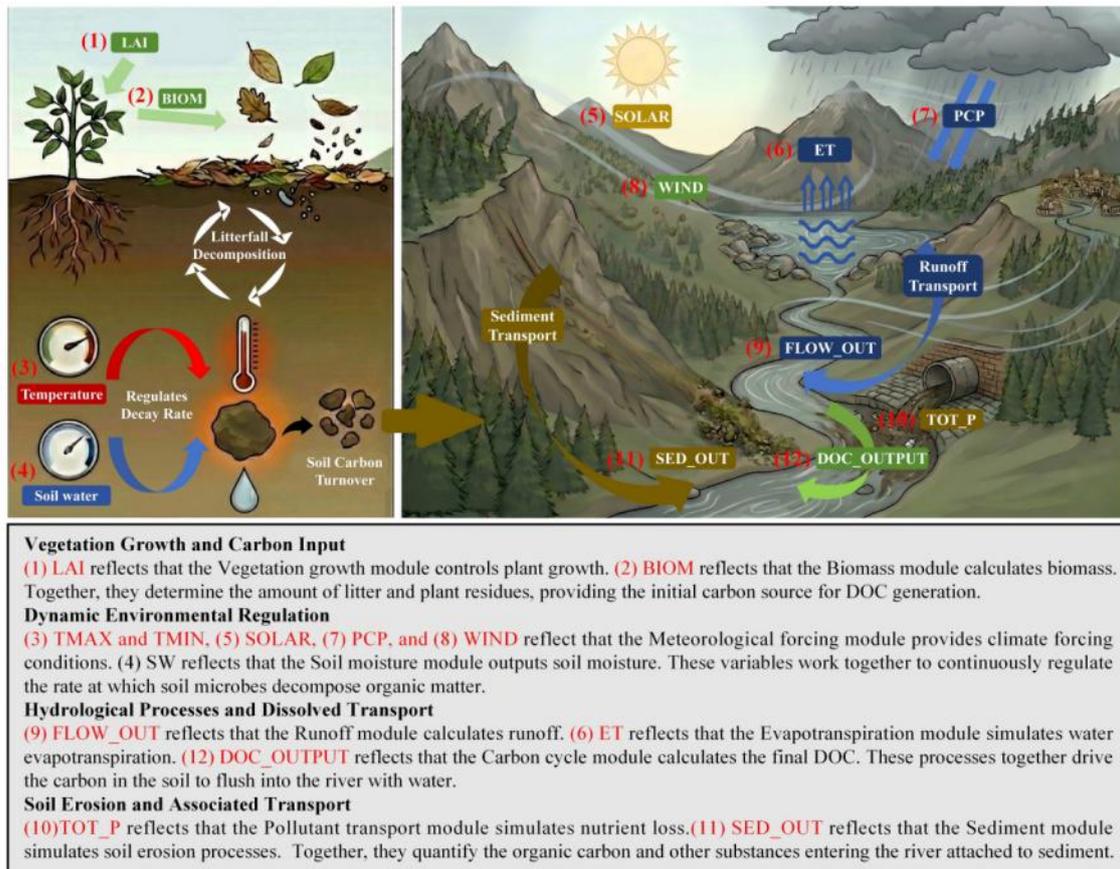*The spatiotemporal evolution of DOC concentrations in the Yalong River Basin provides a direct reflection of how climatic and hydrological conditions drive the carbon cycle. Spatially, DOC concentrations exhibit a progressive increase from the northwest to the southeast, aligning closely with the hydrothermal gradients of the basin. The relatively warm and humid climate in the southeastern region not only*

promotes vegetation productivity but also enhances microbial activity in decomposing organic matter, thereby increasing local carbon release(Davidson and Janssens, 2006). This elevated local supply, combined with the downstream accumulation of streamflow, collectively shapes the high concentration zones observed in the downstream reaches. Temporally, the seasonal distribution of precipitation governs the annual fluctuations of DOC(Blaurock et al., 2025). Concentrated rainfall during the monsoon season triggers a pronounced flushing effect, rapidly leaching soil DOC into the river channel and resulting in significant concentration peaks(Yan et al., 2024).

Whether through spatial biogeochemical control by hydrothermal conditions and vegetation or temporal physical flushing by precipitation, these actual catchment processes driving DOC export are accurately represented within the underlying computing equations of the SWAT-C model. As illustrated in Figure 8, the model characterizes these complex physical and biochemical processes by partitioning them into four core stages, thereby clarifying the specific physical significance of each key feature parameter (as detailed in Table 2) within the model code. This integration of natural phenomena with physical equations demonstrates that the selected feature parameters possess a robust mechanistic foundation.

In the specific computational logic of the model, (1) LAI and (2) BIOM within the vegetation growth and carbon input stage reflect the control of plant growth by the vegetation and biomass modules; these variables determine the quantity of litter and plant residues, providing the initial carbon source for DOC generation(Ji et al., 2022). In the dynamic environmental regulation stage, meteorological forcing conditions including (3) TMAX/TMIN, (5) SOLAR, (7) PCP, and (8) WIND are coupled with (4) SW. Within the underlying code, these are converted into dynamic regulatory factors that continuously modulate the rate of microbial decomposition(Davidson and Janssens, 2006). In the hydrological processes and dissolved transport stage, the water balance calculated by (9) FLOW_OUT and (6) ET, combined with the (12) DOC simulation results, jointly drives the leaching and routing of soil carbon into the river network(Laudon et al., 2011). Furthermore, in the soil erosion and associated

transport stage, (11) SED_OUT and (10) TOT_P simulate soil erosion processes, quantifying the amount of organic carbon that enters the river while attached to sediment particles(Galy et al., 2007). The complete closed loop formed by these four stages ensures that the model accurately captures the comprehensive impacts of environmental changes on the total carbon..

Vegetation Growth and Carbon Input
(1) LAI reflects that the Vegetation growth module controls plant growth. (2) BIOM reflects that the Biomass module calculates biomass. Together, they determine the amount of litter and plant residues, providing the initial carbon source for DOC generation.
Dynamic Environmental Regulation
(3) TMAX and TMIN, (5) SOLAR, (7) PCP, and (8) WIND reflect that the Meteorological forcing module provides climate forcing conditions. (4) SW reflects that the Soil moisture module outputs soil moisture. These variables work together to continuously regulate the rate at which soil microbes decompose organic matter.
Hydrological Processes and Dissolved Transport
(9) FLOW_OUT reflects that the Runoff module calculates runoff. (6) ET reflects that the Evapotranspiration module simulates water evapotranspiration. (12) DOC_OUTPUT reflects that the Carbon cycle module calculates the final DOC. These processes together drive the carbon in the soil to flush into the river with water.
Soil Erosion and Associated Transport
(10)TOT_P reflects that the Pollutant transport module simulates nutrient loss.(11) SED_OUT reflects that the Sediment module simulates soil erosion processes. Together, they quantify the organic carbon and other substances entering the river attached to sediment.

Figure 8. Conceptual diagram of physical mechanisms for DOC in the SWAT-C model.
Numbers (1) through (12) correspond to the key feature parameters and their specific roles in the catchment DOC export process.

Table S4. *Source code locations of the selected SWAT-C process variables*

Data Name	Definition	SWAT-C Module	Locations
FLOW_OUT	Streamflow output	Runoff module	surface_change.f, surfst_h2o_change.f, wattable.f.
SED_OUT	Sediment output	Sediment module	latsed.f,bacteria.f, cfactor_change.f.

			orgncswat_change.f,
DOC_Simulate	*Simulated total DOC*	*Carbon cycle module*	*gw_doc_new.f, substor_change.f, subbasin_change.f.*
BIOM	*Biomass*	*Biomass module*	*plantmod_change.f, grow_change.f*
ET	*Evapotranspiration*	*Evapotranspiration module*	*etpot_change.f, etact_change.f*
TOT_P	*Total phosphorus output*	*Pollutant transport module*	*pestlch.f, pestw.f, pesty.f, nminrl.f, solp_change.f.*
LAI	*Leaf area index*	*Vegetation growth module*	*plantmod_change.f, grow_change.f*
SW	*Soil water content*	*Soil moisture module*	*percmain.f*
PCP, RH, TMAX, TMIN, SR, WIND	*Precipitation, Relative humidity, Max/Min Temperature, Solar radiation, Wind speed*	*Meteorological forcing module*	*pmeas.f, hmeas.f, tmeas.f, smeas.f, wmeas.f,*

[Comment 24] Table S3:What is YLR? Wouldn't it be interesting to study here which model processes are relevant for DOC export?

[Response 24] Following the referee's suggestion, we updated the abbreviation YLR to Yalong River in Table S3 and added a mechanistic analysis of relevant model processes for dissolved organic carbon export alongside a conceptual diagram in Section 4.1, as detailed in our response to comment#23.

[Comment 25] L162:What is meant with "Bayesian optimization was applied"? Which algorithm was applied and on what? Keep in mind that Bayesian optimization is not only used in Machine Learning, it can also be used for model calibration, so it is important to state specifically what has been done.

[Response 25] Following the referee's suggestion, we clarified that Bayesian optimization via the Tree structured Parzen Estimator algorithm was applied strictly to machine learning hyperparameter tuning rather than physical model calibration, and listed the optimal values in Table S2 in the revised manuscript:

"To ensure model robustness and prevent overfitting during this fine-tuning phase, Bayesian optimization utilizing the Tree-structured Parzen Estimator (TPE) algorithm was applied strictly to tune the hyperparameters of the machine learning models. This optimization process targeted the minimization of validation loss by systematically searching the parameter space. The finalized optimal hyperparameter values are presented in Table S2. This established a stable and reliable foundation for the subsequent SHAP-based and residual analyses."

[Comment 26] Structure :I think it would be more intuitive to explain SWAT-C first (currently chapter 2.4) and then the additional layer of SWAT-MDF (currently chapter 2.3)

[Response 26] Following the referee's suggestion, we reorganized the methodology to introduce the SWAT-C model in Section 2.3 before presenting the machine learning SWAT-MDF framework in Section 2.4, and revised the transitional paragraphs for coherent logical progression in the revised manuscript.

[Comment 27] 2.3:This chapter needs to be better explained and better linked with Figure 2. Why are there so many transfers necessary? As this method is not a standard approach in hydrology (yet), try to keep it understandable for readers who are not familiar with it.

[Response 27] We thank the reviewer for this constructive suggestion. We clarified that transfer learning prevents overfitting on sparse observations by pretraining machine learning models on massive physical simulation data to capture mechanisms before fine tuning to correct biases, and explicitly linked this workflow to Figure 2 in the revised manuscript:

"Because deep learning methods are not yet standard approaches for catchment DOC modeling when observed field data are highly limited, a transfer learning strategy was implemented to bridge the gap between abundant physical simulations and sparse field observations. During the transfer process, the models were initially pretrained using massive long term simulation outputs from the SWAT-C model. This

vital pretraining step forces the machine learning algorithms to learn the fundamental physical and hydrological baseline patterns of the basin. Subsequently, the models were finely tuned using the limited observed DOC data. This essential transfer is necessary because it allows the algorithms to correct the systematic biases of the physical model using real world observations while preserving the physical logic learned during pretraining. "

[Comment 28] L227:There is no information on the sensitivity analysis in the supplement material. It is only stated that it was performed, without providing any details. Please provide those details.

[Response 28] We thank the reviewer for pointing out this omission. We added the methodological details of the global sensitivity analysis using the SUFI-2 algorithm within SWAT-CUP and summarized the calibrated parameters alongside their t statistic and p value in Table S3 and Text S1 in the revised manuscript:

"The 2022 release of the SWAT-C model was employed in this study (Yang and Zhang, 2016; Zhang et al., 2013). Model calibration and global sensitivity analysis were conducted using the SUFI-2 algorithm implemented in SWAT-CUP. Model performance during calibration was evaluated using three widely used statistical indicators, including the coefficient of determination (R^2), Nash–Sutcliffe efficiency (NSE), and Kling–Gupta efficiency (KGE). Parameter sensitivity was quantified based on the T-statistic and corresponding P-value generated by the SUFI-2 algorithm. The final parameter values and their sensitivity statistics (T-stat/P-value) for the Yalong River Basin are summarized in Table S3.

Table S3. *Sensitivity analysis and calibration results of SWAT-C parameters.*

Parameter	Change	Calibration range	Calibrated value	Definition	T-Stat/ P-Value
v__CH_K1.sub	Replace	(0,300)	107.250	Effective hydraulic conductivity in tributary channel alluvium	15.72/0.04
v__CH_K2.rte	Replace	(0,500)	253.750	Effective hydraulic conductivity in main channel alluvium	16.27/0.04

......"

[Comment 29] Figure 3: What is meant by the unit DOC/kg? A kilogram of what? Sediment or discharge, or? Is this supposed to quantify the DOC load? Why not work with DOC concentration? I doubt that any DOC-related process can be fitted by working with DOC loads, as it seems mainly discharge driven (Figure 3e).

[Response 29] We thank the reviewer for pointing out this critical issue. We replaced the simulation target from dissolved organic carbon load to concentration, retrained all machine learning models, and updated all related figures and statistics based on the new concentration data in the revised manuscript:

"In contrast, the model's ability to simulate daily DOC was relatively limited, with lower performance (calibration: NSE = 0.57, R² = 0.75, KGE = 0.74; validation: NSE = 0.44, R² = 0.71, KGE = 0.81). Specifically, the model tended to overestimate DOC concentrations during certain high-flow periods, while underestimation was observed under low-flow conditions. This flow-dependent bias contributed to the overall reduction in model performance at the daily scale.
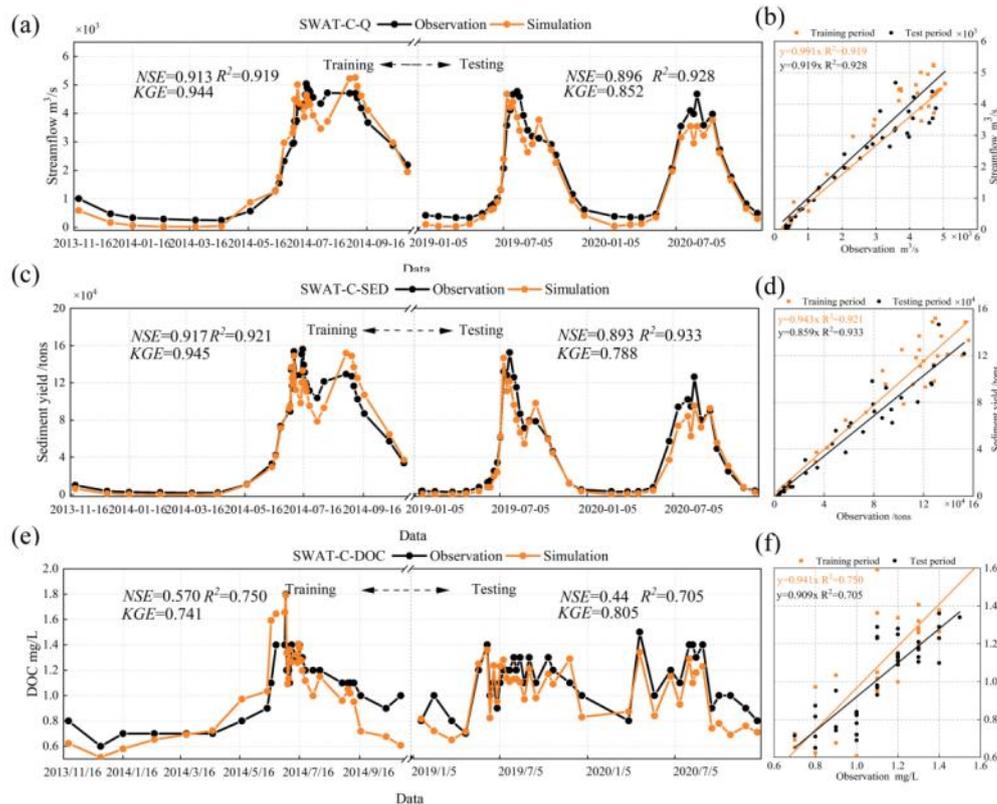

Figure 3. Simulation performance of the SWAT-C model. Line plots (a) streamflow (Q), (c) sediment yield (SED), and (e) dissolved organic carbon (DOC). Scatter plots (b), (d), and (f) show the correlations between observed and simulated values."

[Comment 30] Figure5 and 6:Why is the LAI the least important in Figure 5 and the

most important in Figure 6b?.

[Response 30] We thank the reviewer for this observation. We clarified that Figure 5 measures direct driving effects on predicted concentration via feature importance, whereas Figure 6b measures structural uncertainty via residual attribution analysis, identifying the LAI vegetation growth module as the largest source of simulation error in the revised manuscript:

"The vegetation growth module, represented by LAI, exhibited low importance in the global SHAP analysis but emerged as a dominant factor in the residual analysis. To investigate this discrepancy, a single feature SHAP analysis was conducted for LAI (Text S4 in the Supplementary Materials), revealing a negative effect on the model output. Because LAI possesses relatively low temporal variability and provides a stable numerical contribution over time, its global SHAP importance remained limited in determining the absolute dissolved organic carbon concentration. However, the exceptionally high residual contribution associated with LAI indicates severe structural uncertainties within the vegetation growth module. This explicit contrast highlights that while vegetation dynamics do not numerically dominate the prediction, the underlying physical mechanism representation requires targeted refinement to improve the overall simulation reliability."