Response to Referee #2

[**Comment 1**] This manuscript proposes a hybrid framework (SWAT-MDF) that combines the physically based SWAT-C model with machine-learning approaches and SHAP-based interpretation to diagnose module-level uncertainty in dissolved organic carbon (DOC) simulations. The topic is relevant, as DOC modelling remains challenging and improved diagnostic tools for process-based models are indeed needed.

However, in its current form, the study does not convincingly demonstrate sufficient methodological rigor nor provide results that are practically useful or actionable. While the conceptual idea is potentially interesting, the implementation raises substantial concerns regarding conceptual consistency, experimental design, interpretability, validation strategy, and scientific contribution. Consequently, several conclusions appear overstated relative to the evidence presented.

[**Response 1**] We thank the reviewer for the detailed evaluation and recognizing the relevance of the framework. We restructured the methodology, incorporated four years of daily runoff observations for rigorous model calibration, and added a conceptual diagram explicitly linking input variables to physical carbon processes. Furthermore, we emphasized applying polynomial feature expansion and ridge regression to the model residuals, rather than relying solely on SHAP for module diagnosis, to prove our machine learning diagnosis identifies actual physical flaws instead of mere statistical patterns. Please see below our responses to each comment.

Major comments

[**Comment 2**] T1. Conceptual framework and interpretation

The central claim of the manuscript is that machine learning can be used to diagnose structural uncertainty in the SWAT-C model. However, the proposed workflow effectively trains a machine-learning surrogate model using SWAT outputs as predictors and observed DOC as targets (Section 2.3). This creates several conceptual issues.

First, the ML model learns statistical relationships between SWAT outputs and observations, rather than diagnosing model structure itself. The authors assume that selected SWAT-C outputs represent individual model modules, and therefore that higher feature importance in the ML model implies higher uncertainty or relevance of the corresponding module. This logical step is not sufficiently justified.

For example:

Simulated DOC (DOC_Simulate) is a final model output resulting from multiple interacting modules and processes. It cannot uniquely represent the carbon cycle module alone.

The manuscript concludes that the vegetation (LAI) module introduces structural uncertainty in DOC simulations. However, LAI is an independent state variable that can be calibrated and validated separately. High statistical importance or residual contribution does not necessarily imply structural deficiencies in that module.

Therefore, SHAP importance reflects feature usefulness for the ML predictor, not causal deficiencies of SWAT modules. The manuscript does not provide theoretical or empirical evidence demonstrating that feature attribution can be translated into module-level structural errors.

A related issue arises in the SHAP results. The analysis identifies DOC_Simulate as the dominant feature. This outcome is expected because DOC_Simulate is already the SWAT-simulated DOC output and therefore strongly correlated with observations. This result is essentially tautological rather than diagnostic. In fact, it may indicate that SWAT-C already captures a substantial portion of DOC variability. Consequently, the conclusion that the carbon cycle module is the primary uncertainty source is not convincingly supported.

Finally, the study concludes that future improvements should focus on the carbon cycle module, vegetation module, and relative humidity inputs. These recommendations remain generic and largely consistent with existing DOC modelling literature. No specific parameters, process representations, or structural modifications are identified. As such, the framework currently provides limited actionable guidance for model development.

[Response 2] We thank the reviewer for the detailed critique regarding the conceptual framework. We highlighted the residual analysis using polynomial feature expansion and ridge regression to quantify how intermediate physical state variables drive systematic simulation errors, and introduced a conceptual diagram linking input variables to carbon processes to demonstrate the framework identifies actual structural flaws rather than mere statistical patterns. Please see below our detailed responses to each specific point.

1.Regarding the overall conceptual framework and residual analysis

We highlighted applying polynomial feature expansion and ridge regression to model residuals to explicitly quantify how intermediate physical state variables drive

systematic simulation errors, and incorporated a new discussion in Section 4.1 with an added conceptual diagram linking input variables to carbon processes to diagnose actual structural flaws rather than mere statistical patterns in the revised manuscript:

"Residuals, which are defined as the differences between predicted and observed values, were further analyzed to assess the model bias and predictive uncertainty. To quantify the influence of input features on prediction errors, residuals were computed for each sample, and polynomial feature expansion was applied to generate nonlinear interaction terms. Ridge regression was subsequently used to explore the relationship between these terms and the residuals (Santos Nobre and Da Motta Singer, 2007; Tyagi et al., 2022). The absolute values of the regression coefficients served as indicators of each feature's contribution to the residuals (Eq. 1). This allowed for the quantification of both individual feature effects and interaction-driven contributions to model errors. To evaluate the robustness of these estimates, a bootstrapping approach was used to repeatedly resample the training dataset (Hongyi Li and Maddala, 1996), and the mean and standard deviation of each feature's contribution rate were computed to characterize uncertainty. Two residual analysis strategies were adopted: (1) Interaction-based: capturing nonlinear feature interactions via polynomial expansion and ridge regression; (2) Single-feature: assessing individual feature contributions without interactions.

$$C_{\mathrm{j}} = \frac{\left|\beta_j\right|}{\sum_{\mathrm{i}=1}^{n}\left|\beta_{\mathrm{i}}\right|} \times 100\% \tag{1}$$

where C_j denotes the contribution rate of the j-th feature to the residuals, expressed as a percentage; β_j denotes the ridge regression coefficient corresponding to the j-th feature; n denotes the total number of input features.

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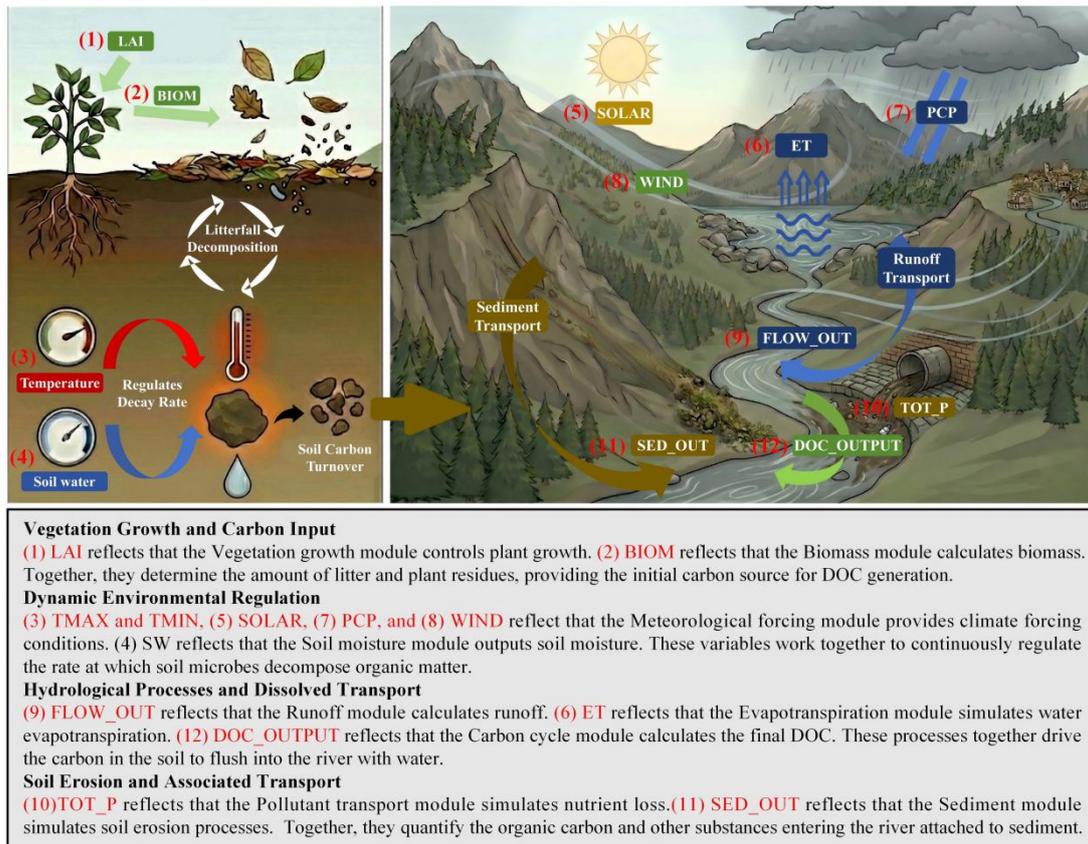
Figure 8. Conceptual diagram of physical mechanisms for DOC in the SWAT-C model. Numbers (1) through (12) correspond to the key feature parameters and their specific roles in the catchment DOC export process."

2. Regarding the representation of simulated dissolved organic carbon

We agree that DOC_Simulate is a final output shaped by multiple interacting processes and cannot uniquely represent the carbon cycle module alone. For this reason, we further conducted pairwise interaction analysis in the residual framework. By generating interaction terms among all features, we decomposed this integrated output into identifiable structural components and thereby diagnosed more specific process-related weaknesses:

"3.4.2. Residual Analysis of Input Features

Residual analysis was further conducted to quantify the contribution of input features to DOC prediction residuals within the coupled model. The three interaction terms with the highest contributions were identified as key factors. As illustrated in the residual interaction analysis (Fig. 7a), the dominant contributors were DOC_Simulate-TMIN, SW-RH, LAI with contribution rates of 2.83%, 2.21%, 2.03% respectively. Notably, the identified interaction terms involve variables related to

climatic and hydrological conditions, indicating that environmental factors jointly influence the residual patterns of DOC simulations. Additionally, LAI emerged as an important individual contributor, suggesting that vegetation conditions play a notable role in shaping the distribution of model residuals.
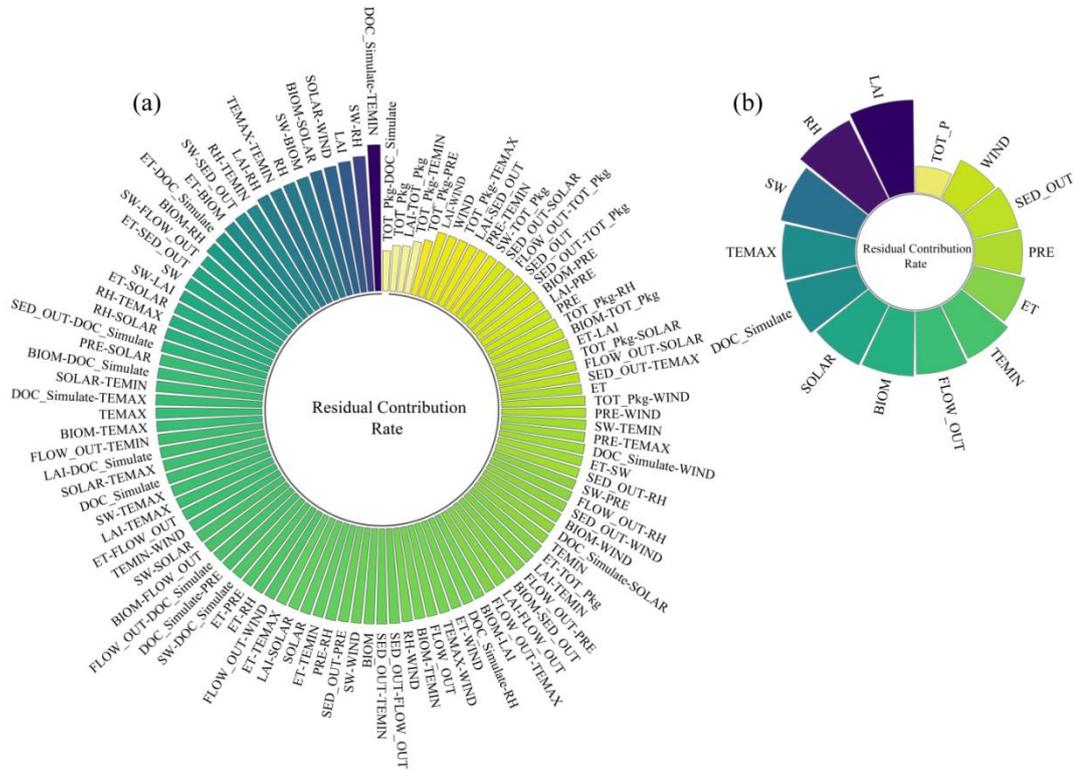
Figure 7. Residual analysis of characteristic parameters. *(a) presents the detailed contribution rates for all parameter interactions, while (b) summarizes the primary individual contributions of key parameters. Circular bar plots illustrating the residual contribution rates of input parameters and their pairwise interactions influencing model DOC predictions. "*

3.Regarding the calibration of leaf area index and structural uncertainty

We agree that LAI is an independent state variable and can be calibrated. To reduce parameter uncertainty, we first calibrated the key physical processes, including runoff, sediment, and DOC, and further clarified this issue in the revised Discussion section. Under this optimized parameter space, the dominant residual contribution from LAI dynamics points to structural limitations in the vegetation module. Following the referee's suggestion, we revised the relevant expression in the revised manuscript as:

"3.1. Model performance of SWAT-C Model in DOC simulations

The first component of the coupled model, the SWAT-C model, achieved excellent

performance in simulating daily runoff and sediment for the Yalong River Basin (Fig. 3). During the calibration period, the SWAT-C model achieved high accuracy for daily runoff (calibration: NSE = 0.91, R^2 = 0.92, KGE = 0.94; validation: NSE = 0.90, R^2 = 0.93, KGE = 0.85). For sediment, both calibration and validation phases yielded excellent results (calibration: NSE = 0.92, R^2 = 0.92, KGE = 0.95; validation: NSE = 0.89, R^2 = 0.93, KGE = 0.79). In contrast, the model's ability to simulate daily DOC was relatively limited, with lower performance (calibration: NSE = 0.57, R^2 = 0.75, KGE = 0.74; validation: NSE = 0.44, R^2 = 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.
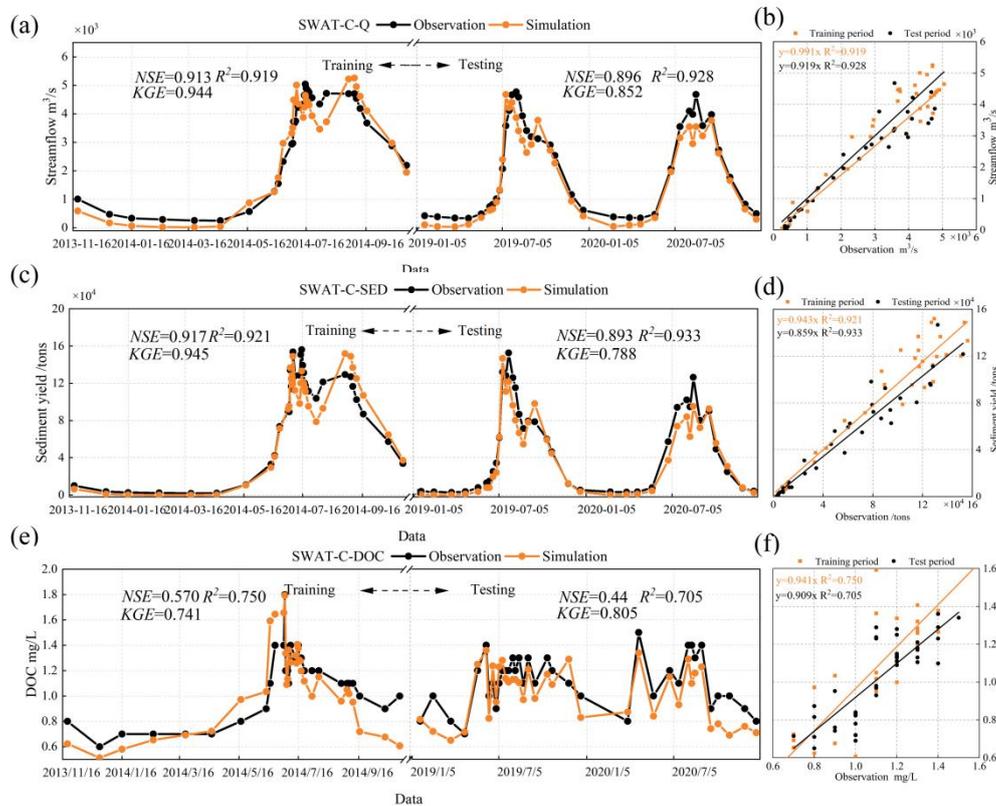
***Figure 3. Simulation performance of the SWAT-C model.** Line plots (a) streamflow (Q), (c) sediment yield (SED), and (e) dissolved organic carbon (DOC). Scatter plots (b), (d), and (f) show the correlations between observed and simulated values.*

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."

4. Regarding the interpretation of SHAP results

We agree that the high global SHAP importance of DOC_Simulate is expected and mainly reflects the baseline predictive information provided by SWAT-C, rather than a structural deficiency. Our inference regarding carbon-module uncertainty is therefore not based on SHAP ranking alone, but on the combined evidence from global SHAP interpretation and residual interaction analysis. Following the referee's suggestion, this point has been clarified in our response to Point 2 above.

5. Regarding actionable guidance and specific structural modifications

We appreciate this helpful comment. We agree that identifying broad modules alone is insufficient for guiding model development. In the revised manuscript, we therefore expanded the Discussion section to link each selected feature to its corresponding DOC-related physical process, and strengthened the reference to Table S4, which provides the relevant subroutines. The added discussion in Section 4.1 is presented in our response to Point 1. Following the referee's suggestion, we revised the relevant text as:

*"**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.
DOC_Simulate	Simulated total DOC	Carbon cycle module	orgncswat_change.f,

			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 2] 2. Data DOC calibration and validation/testing relies on only 34 + 38 samples. This is critically insufficient for training deep learning models although the authors used transfer learning validation. Given such sparse observations, ML results are likely dominated by SWAT simulations used in pretraining. Although the authors acknowledge data scarcity but proceeds with complex ML architectures that require large datasets. This contradiction undermines the credibility of the analysis.

[Response 2] We thank the referee for this important comment. We agree that the limited DOC observations are insufficient for training a purely data-driven deep learning model. However, our framework is not intended as a standalone predictor. The pretraining stage is designed to preserve the physically based structure of SWAT-C, while the limited observations are used only for fine-tuning to reveal systematic discrepancies between model simulations and real-world processes. We also added four years of observed runoff data to improve SWAT-C calibration and strengthen the physical basis of the analysis:

"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. Text S3 in the Supplementary Materials details the principles of the selected ML methods and the procedures for Bayesian-based hyperparameter optimization.

To further demonstrate the robustness of the hydrological processes represented in the SWAT-C model, daily observed streamflow during 2013, 2014, 2019, and 2020 was additionally used for model calibration and evaluation. The model achieved satisfactory performance with NSE= 0.67 and R² = 0.67, indicating that the model reasonably captures the daily runoff dynamics of the basin. The comparison between observed and simulated streamflow is provided in Figure S3 in the Supplementary Materials.
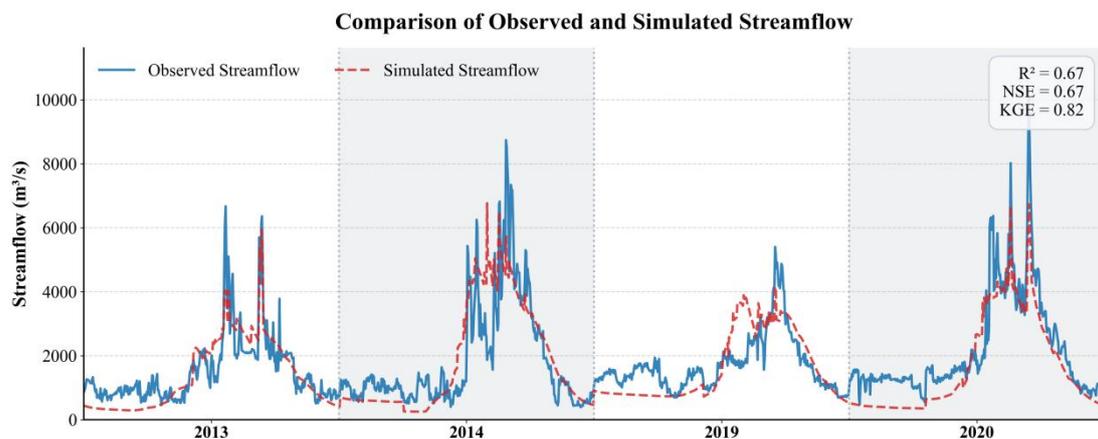

Figure S3. Comparison between observed and simulated daily streamflow used for calibration of the SWAT-C model. "

Minor comments

[Comment 3] Line 87: SWAT-MDF is introduced without providing its full name at first mention.

[Response 3] We thank the reviewer for pointing out this omission. We have corrected it in the revised manuscript by providing the full name, Soil and Water

Assessment Tool Model Diagnostic Framework (SWAT-MDF), at its first occurrence. The corresponding revised text is as follows: *"The SWAT-MDF (Soil and Water Assessment Tool Model Diagnostic Framework) quantitatively diagnoses module-level sources of uncertainty in SWAT-C and evaluates their relative contributions under complex environmental conditions."*

[Comment 4] Line 129: The sentence appears incomplete and should be revised.

[Response 4] We appreciate your careful reading. We agree that the original sentence was incomplete and lacked a clear connection to the surrounding context. To improve the overall clarity and coherence of the paragraph, we have removed this sentence entirely in the revised manuscript.

[Comment 5] Line 148: Terminology alternates between parameters, outputs, and features without clear definitions; consistent terminology is needed.

[Response 5] We thank the reviewer for pointing out this terminology inconsistency. We have standardized the terminology throughout the revised manuscript. Specifically, parameters refers to the calibration settings of SWAT-C, simulated variables to the physical outputs of SWAT-C, and feature variables to these outputs when used as inputs to the data-driven module. The corresponding paragraph and Table 2 have been revised accordingly. The revised text is as follows:

"The first component (Fig. 2a), SWAT-C model, provides the physical and mechanistic foundation of the diagnostic framework (Section 2.3.1). Its simulations are driven by spatial datasets such as digital elevation model (DEM), soil type, land use and slope, as well as meteorological variables including precipitation, relative humidity, maximum and minimum temperature, wind speed and solar radiation. Upon model calibration (More details are provided in Text S1 in the Supplementary Materials.), a subset of key simulated variables that govern DOC dynamics within the watershed was identified. "

[Comment 6] Table 2: The criteria for selecting outputs as representatives of specific modules are unclear. For example, justification is needed for why TOT_P represents the pollutant transport module.

[Response 6] We thank the reviewer for this helpful comment. We have clarified that the selected variables were chosen based on their direct relevance to DOC generation

and transport in SWAT-C. Specifically, TOT_P was used to represent the pollutant transport module because it is closely linked to sediment and surface-runoff transport pathways and also reflects biogeochemical processes affecting DOC production. To improve clarity, we added a dedicated explanation of the selection rationale for all variables in the methodology section. The revised text is as follows:

"These simulated variables, each linked to specific physical modules in the SWAT-C structure, are used as feature variables for the second component of the framework. Tab. 2 provides detailed definitions of these feature variables and their associated modules within the SWAT-C model. The selection of these feature variables is strictly based on their physical connections to carbon cycling. Meteorological forcings act as fundamental climatic drivers. Streamflow, sediment, soil moisture, and evapotranspiration represent the physical water and mass transport pathways. Biomass and leaf area index govern the biological production and storage of carbon. Total phosphorus is explicitly chosen to represent the pollutant transport module because it shares identical physical transport pathways with organic carbon and acts as a limiting nutrient for biological carbon generation. This comprehensive selection ensures the diagnostic framework evaluates the complete chain of internal processes governing watershed carbon dynamics."

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.

[Comment 7] Figure 4: Panels (a) and (b) are too small to read and should be enlarged for clarity.

[Response 7] We thank the reviewer for pointing out the legibility issue with Figure 4. We agree that panels (a) and (b) were too small for proper visual inspection. In the revised manuscript, we have significantly enlarged these two panels and optimized the overall layout of the figure.