Reply to Reviewers' comments (Reviewer#2)

Ref: Manuscript ID egusphere-2025-4778

Title: Disentangling the Key Drivers of Water Balance in Central Asia's Lake Balkhash: A Relative Contribution Assessment (Original title: Revealing the Driving Factors of Water Balance in Lake Balkhash Through Integrated Attribution Modeling)

Dear Reviewer,

We would like to express our sincere gratitude for your constructive and insightful comments on our manuscript. We appreciate the time and effort you have dedicated to reviewing our work. We have carefully considered all your suggestions. Below, we provide a point-by-point response to your comments. The reviewer's comments are highlighted in red, and our responses are highlighted in black.

Major comments:

1. Inappropriate terminology regarding the PIML.

The manuscript characterizes the proposed model as Physics-Informed Machine Learning (PIML); however, it looks more like a ML-corrected SEGSWAT+ to me. In this study, the physics-based model (SEGSWAT+) run independently, and a ML model is subsequently trained to predict the discrepancy between the simulated outputs and observations. While this strategy can improve predictive skill, it does not incorporate physical laws, constraints, or governing equations into the learning process itself. As such, the ML component operates as a statistical correction to the physics model rather than being informed by physics during model training or optimization.

Under commonly used definitions, PIML frameworks require explicit physical constraints to be embedded within the model architecture, loss function, or parameter evolution (see Raissi et al., 2019; Shen et al., 2023). The proposed method would therefore be more accurately described as ML-corrected SEGSWAT+ or a hybrid model rather than a PIML. I would suggest the authors change the terminology in order to avoid conceptual ambiguity and ensure consistency with established definitions in the literature.

Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. J. Comput. Phys., 378, 686–707. doi: 10.1016/j.jcp.2018.10.045

Shen, C., Appling, A. P., Gentine, P., Bandai, T., Gupta, H., Tartakovsky, A., ...Lawson, K. (2023). Differentiable modelling to unify machine learning and physical models for geosciences. Nat. Rev. Earth Environ., 4, 552–567. doi: 10.1038/s43017-023-00450-9

Response: We greatly appreciate the reviewer's crucial conceptual clarification. We fully agree with your assessment that our initial use of the term "Physics-Informed Machine Learning (PIML)" did not strictly adhere to the core idea of the field (as defined by Raissi et al., 2019; Shen et al., 2023), which is to directly embed physical laws during machine learning training. Our model architecture does indeed better fit the description of a "hybrid model" or "physical model for machine learning error correction."

To ensure accuracy in terminology and clarity of concept, we have adopted your suggestion and made systematic revisions throughout the paper: Terminology Correction:

(1) We have replaced the term "PIML" throughout the paper with the more accurate "hybrid hydrological model" or "a framework that couples the process-based... model with a ML error-correction module."

(2) Updated Method Description: In section 2.3.1, "Hybrid Hydrological Reconstruction Model," we have revised the model's structure, explicitly stating that it is a two-stage hybrid modeling strategy, rather than an end-to-end, physically constrained machine learning model. We emphasize that the advantage of this approach lies in leveraging a physical model to provide a physically consistent benchmark simulation, which is then learned and corrected for by the ML model to correct for systematic residuals.

We believe this revision makes our methodological description more rigorous and consistent with current academic definitions. Thank you again for your accurate correction.

.

1. Research gap

Line 50-60: The research gap is not clearly explained. Previous studies have already developed multiple models to quantify the contributions of different drivers to lake water balance changes. For example, Yu et al. (2025) developed a distributed Geomorphology-Based Hydrological Model (GBHM) to quantify the contributions of multiple drivers. What is the key difference or advancement of your approach compared to GBHM in terms of the study objective (i.e., driver attribution)? The authors claim that previous models did not “integrate their findings with the lake’s terminal water balance,” but this statement is vague. What does the “terminal water balance” mean exactly? Does this imply that previous studies did not directly simulate lake water levels or storage changes?

In addition, the authors state that data scarcity, particularly limited lake inflow observations, is a major challenge for existing hydrological models. It is unclear why this limitation would affect physics-based hydrological models and machine-learning models, but not the proposed hybrid model. If the uncertainty arises from insufficient data for model calibration, this limitation would appear to be a general issue for all modeling approaches rather than one unique to existing models.

It's very important (perhaps the most important) to clearly and explicitly articulate the specific research gap, the limitations of previous studies, and how the proposed approach meaningfully advances beyond existing models.

Response: Thank you for pointing out the shortcomings in our description of the research gaps. Your feedback prompted us to re-examine and more clearly articulate the core contributions of this study. We have made significant revisions to the introduction to explicitly answer your questions.

Specific revisions and explanations are as follows:

Identifying the Research Gap (Differences from Existing Research): We now clearly identify two key gaps.

(1) The first gap is methodological: We acknowledge that data sparsity is a common challenge for all models. However, we highlight the unique advantages of hybrid models in addressing the tradeoff between uncertainty and data sparsity. In the introduction to the revised manuscript, we stated: "The specific advantage of the hybrid approach lies in mitigating this 'uncertainty vs. data scarcity' trade-off. By using a physics-based model to simulate the fundamental hydrological processes and then employing ML solely to learn the residuals, this approach enforces physical constraints while effectively correcting the structural biases of the physical model, improving accuracy beyond what traditional calibration can achieve with limited data."

(2) The second gap concerns the "broken chain" problem at the research scale: Your question about "terminal water balance" is very apt. We mean that many studies (such as Yu et al., 2025, which you mentioned) quantify the driving factors at the watershed outlet, but fail to directly and quantitatively transfer and link the contributions of these upstream driving factors (such as a reduction in glacial meltwater at a specific volume vs. an increase in agricultural water diversion) to the lake's own water volume and level changes. In our revised introduction, we clarified this point: "...existing studies typically focus on decomposing streamflow changes at the catchment outlet but fail to explicitly link these catchment-scale drivers to the lake water storage volume and water level. ... This disconnect prevents a direct quantitative explanation of how specific upstream drivers... translate into the observed vertical fluctuations of the lake itself, which is the

ultimate metric of ecological health." Our study, through a three-stage framework, achieves an end-to-end quantitative link from "separation of watershed runoff drivers" to "separation of lake water volume change drivers," a key advancement compared to previous research.

We believe that, with these revisions, the positioning and innovative aspects of this study have been more clearly and powerfully articulated.

3. Calibration details

Section 2.3.1 requires additional detail regarding the model calibration strategy. Specifically:

(1) The calibration (training), validation, and testing periods should be clearly specified here. (2) Based on Figure 3, the overall strategy appears to involve pre-calibrating SEGSWAT+ using gauge station observations, followed by training the machine-learning model to correct the residuals. If this interpretation is correct, it should be stated explicitly in Section 2.3.1 to avoid confusion. (3) What's the hyper parameter selection strategy for each ML/DL model (e.g., number of layers for ANN, sequential length for LSTM)? (4) A table summarizing the SEGSWAT+ parameters used for calibration, as well as the machine-learning hyperparameters, is essential. This table could be placed in the appendix. (5) A comparative table reporting NSE, KGE, PBIAS, and R^2 for the raw SEGSWAT+ outputs and the final ML-corrected model across the calibration (training), validation, and testing periods should be provided to clearly demonstrate the performance improvement achieved by the ML correction. (6) Section 2.3.4 could be merged into Section 2.3.1. Dedicating an entire section to explaining KGE, NSE, and PBIAS is unnecessary, as these metrics are widely used and well understood in hydrological modeling. (7) Multiple evaluation metrics are used in this study. If these metrics yield conflicting assessments, how is the optimal parameter set selected?

Response: Thank you for your specific and crucial suggestions regarding the details of model calibration. A fully transparent and repeatable calibration process is the cornerstone of research. We have comprehensively supplemented and reorganized the methods section based on your suggestions.

Specific modifications are as follows:

(1) Clearly define the time periods: In the caption of Figure 6, we clearly state: "The shaded gray background indicates the calibration period, while the unshaded area represents the validation period." Additionally, in section "2.3.1 Hybrid...", we have supplemented the explanation of the dataset partitioning for the machine learning part: "The dataset for each period was split into training (70%) and validation (30%)

subsets."

(2) Clearly define the two-stage workflow: Your understanding of our workflow is entirely correct. We have clearly outlined this two-stage calibration strategy at the beginning of section “2.3.1 Hybrid…”: “The workflow proceeds in two distinct stages… First, the SEGSWAT+ model was independently calibrated using observed streamflow… Subsequently, the residuals… were calculated. A suite of ML algorithms was then trained to predict these residuals…”

(3) Hyperparameter selection strategy: We supplemented the hyperparameter optimization method in section “2.3.1 Hybrid…”: “Hyperparameters for each model were optimized using a grid search strategy (details in Appendix Table B2).”

Table B2. Hyperparameter optimization ranges and selected values for the machine learning models

Model	Hyperparameter	Search Range	Optimal Value
ANN	Hidden Layers	[1, 2, 3]	2
	Neurons per Layer	[16, 32, 64, 128]	64
	Learning Rate	[0.001, 0.01, 0.1]	0.01
	Activation Function	[ReLU, Tanh, Sigmoid]	ReLU
LSTM	Hidden Units	[32, 64, 128, 256]	128
	Lookback Window	[5, 10, 15, 30] days	15
	Dropout Rate	[0.1, 0.2, 0.3, 0.5]	0.2
	Epochs	[50, 100, 200]	100
Random Forest	n_estimators (Trees)	[100, 300, 500, 1000]	500
	Max Depth	[10, 20, 30, None]	20
	Min Samples Split	[2, 5, 10]	5
XGBoost	Learning Rate (eta)	[0.01, 0.05, 0.1, 0.3]	0.05
	Max Depth	[3, 5, 7, 9]	7
	n_estimators	[100, 500, 1000]	500
	Subsample	[0.6, 0.8, 1.0]	0.8

(4) New parameter summary table: We added Appendix Table B2 to the appendix, which details the hyperparameter search range and the finally selected optimal values for each machine learning model.

(5) New performance comparison table: To quantitatively demonstrate the superiority of our hybrid method, we added Appendix Table B3 to the appendix. This table provides a detailed comparison of the performance metrics (KGE, NSE, PBIAS) of the original SEGSWAT+ model and the final hybrid model during the calibration and validation periods at major hydrological stations, clearly demonstrating the significant performance improvement brought about by ML error correction.

Table B3. Performance comparison of SEGSWAT+ (Raw) and the Hybrid Model (Corrected) across calibration and validation periods

River	Station	Period	Metric	SEGSWAT+ (Raw)	Hybrid Model (Corrected)
Ili	Ushzharma	Calibration	KGE	0.68	0.89
			NSE	0.72	0.93
			PBIAS(%)	-9.5	3.2
		Validation	KGE	0.65	0.85
			NSE	0.68	0.88
			PBIAS(%)	-16.8	5.1
Karatal	Ushtobe	Calibration	KGE	0.74	0.89
			NSE	0.76	0.91
			PBIAS(%)	11.2	6.4
		Validation	KGE	0.71	0.86
			NSE	0.72	0.85
			PBIAS(%)	18.5	7.5
Aksu	Chann	Calibration	KGE	0.66	0.83
			NSE	0.64	0.84
			PBIAS(%)	-9.3	-2.8
		Validation	KGE	0.62	0.80
			NSE	0.60	0.78
			PBIAS(%)	-13.5	-3.4
Lepsy	Lepsinsk	Calibration	KGE	0.70	0.82
			NSE	0.71	0.84
			PBIAS(%)	9.8	-5.1

		Validation	KGE	0.68	0.80
			NSE	0.67	0.77
			PBIAS(%)	11.5	-6.2
Ayaguz	Ayaguz	Calibration	KGE	0.63	0.89
			NSE	0.61	0.88
			PBIAS(%)	-15.4	-0.5
		Validation	KGE	0.71	0.86
			NSE	0.68	0.83
			PBIAS(%)	-8.45	-1.8

(6) Binding Section: This is an excellent suggestion. We have integrated the content of the original "2.3.4 Model Evaluation and Uncertainty Metrics" into the "2.3.1 Hybrid…" section, making the description of the methodology more concise and fluent.

(7) Clarifying the Criteria for metric selection: In the "2.3.1 Hybrid…" section, we have added the decision criteria when evaluation metrics conflict: "In cases where metrics yielded conflicting assessments, the KGE was prioritized as the primary selection criterion due to its balanced decomposition of correlation, bias, and variability errors, with PBIAS acting as a constraint to ensure water balance closure."

Minor comments:

(1) Line 27-31: These texts do not explain the environmental issue well. "This balance is under pressure.." on which direction? Increasing or decreasing water storage… Please state the issue clearly. Consider use simple sentences: "Decreasing water storage has become a widespread issue for these lakes, posing a significant threat to their ecological health (reference). The decline in water storage is driven by two primary factors: climate change and human activities. (reference)."

Response: Thank you for your suggestions, which made the statement of the problem more direct and powerful. We have adopted your wording and revised it to: "However, decreasing water storage has become a widespread issue for these lakes, posing a significant threat to their ecological health (Li et al., 2025). This decline is primarily driven by two concurrent forces: ..."

(2) Line 41: omit "Lake Balkhash has no outlet".

Response: Following your suggestion, the redundant information "Lake Balkhash has no outlet” has been removed.

(3) Line 43: “It signals a long-term depletion of solid water reserves”. What does that mean? Do you mean the increasing evaporation outweighs the glacier melt? If yes, it is important to add references to support your statement. Consider “While increasing glacier melt can temporarily raise inflow, the associated increase in evaporation outweighs this effect and leads to overall water depletion.”

Response: Thank you for pointing out the ambiguity here. We wanted to express the non-renewable depletion of solid water reserves by glaciers. We have revised it to: “While increasing glacier melt can temporarily raise inflow, it leads to the irreversible depletion of solid water reserves. This continuous loss of ice storage implies that the current meltwater increase is transient, and future water availability will be threatened as the glacial volume diminishes.”

(4) Line 67: Swap “To achieve this” with “Specifically”

Response: Revised.

(5) Figure 1: Consider remove political borders and just focus on watershed boundaries.

Response: Taking into account the reviewers' suggestions, we assessed the impact of removing the border lines on the communication of map information. Since the Lake Balkhash basin is a transboundary basin (Kazakhstan and China), the border lines are of significant reference value for understanding potential transboundary water resource management issues in the region. Therefore, we prefer to retain the border lines and hope for your understanding.

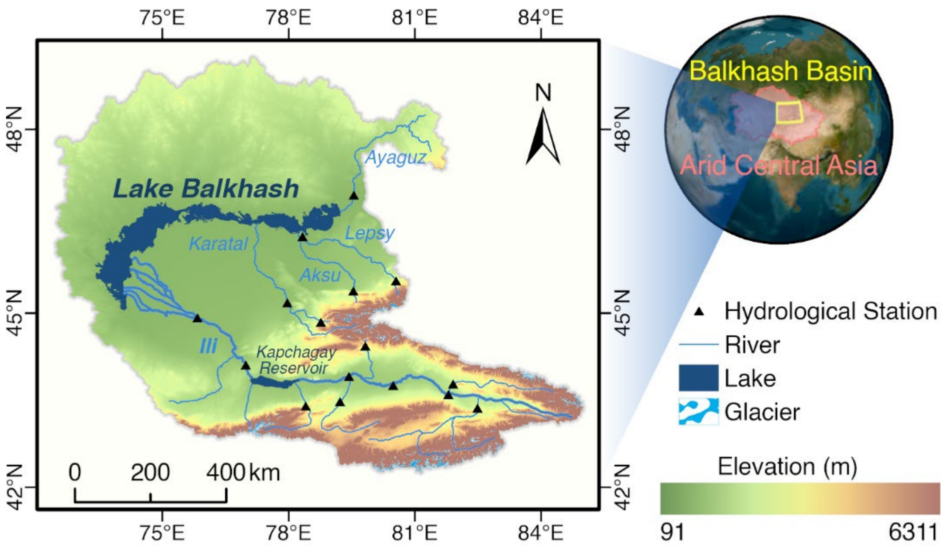

Figure 1. Geographic location of the study area.

(6) Table 1: I appreciate this style! Just a minor suggestion: consider place references in "Source" column instead of simply saying it's from Zenodo.

Response: Your suggestion is very correct. We have revised the content of the "Source" column in Table 2 to the full reference format (e.g., Harris, 2024).

Table 2. Summary of datasets used in this study

Dataset	Key Variables	Spatial Resolution	Temporal Coverage	Source (Reference)
Copernicus GLO-90 DEM	Elevation	90 m	Static	European Space Agency (2019)
DSOLMap	Bulk density, hydraulic conductivity, available water capacity	250 m	Static	Lopez-Ballesteros et al. (2023)
GLC_FCS30D	Land cover classes (35 subcategories)	30 m	1985–2022	Google Earth Engine(Zhang et al., 2024)
Randolph Glacier Inventory (RGI v7.0)	Glacier outlines, attributes	Vector	Target year: 2000 (varies by region)	RGI Consortium (2023)
SWORD v15	River reaches, nodes, hydrological networks, lake boundaries	~10 km reaches, 200 m nodes	Static	(Altenau et al. (2021)
Glacier mass loss	Glacier elevation change rates (dh/dt)	100 m	2000–2019	Hugonnet et al. (2021)
CRU JRA v3.0	Temperature, precipitation, wind speed, vapor pressure, etc.	0.5° (downscaled to 0.05°)	1901–2024 (daily)	Harris (2024)
TerraClimate	Max/min temperature, precipitation, solar radiation, vapor pressure deficit	1/24°	1958–2024 (monthly)	Abatzoglou et al. (2018)
NEX-GDDP-CMIP6	Daily temperature (max/min), precipitation	0.25°	2015–2100 (Daily)	Thrasher et al. (2022)
Observations	Discharge, water level	Point	1931–2024 (monthly)	NCDC (2024); Duan et al. (2020)

(7) Figure2: Spell AAF in the caption. Figure and table captions should be self-explanatory. All acronyms must be fully spelled out in the captions, even if they have already been defined in the main text. Also, check the font style of the caption.

Response: We have spelled all abbreviations (such as HADF) in full in the figure captions and checked and standardized the font of the captions.

(8) Line 140: "The SWAT model and its improved versions are widely used in hydrological simulation processes.", such as? Add references of the original literature of the SWAT model and other publications of the model application.

Response: Based on your suggestion, we have added references to the application of the SWAT model in the text, such as (Forgrave et al., 2024; Ho et al., 2025; Sánchez-Gómez et al., 2025).

(9) Line 141: Swap "Iteration" with "variant".

Response: "Iteration" has been changed to the more accurate "variant".

(10) Figure 3: Need a higher-resolution figure, the Q_final figure is blurred. Also, this flow diagram is not explained well in the main text. What's the relationship between Qsim and Qres. What's Qres? Is it the discharge into reservoir or the residual of simulated discharge? Consider explain this figure component by component in section 2.3.1.

Response: We have replaced it with a higher resolution image. In section "2.3.1 Hybrid…", we have added a detailed explanation of each component of the flowchart, clarifying the relationship between Q_{sim} (physical model simulation of runoff), Q_{obs} (observed runoff), and Q_{res} ($Q_{obs} - Q_{sim}$, i.e., residuals).

(11) Line 189: Is "A" static? Or it was calculated by a hypsometric curve between Area and Storage?

Response: In the section "2.3.3 Lake System Response Linkage", we clarified that the lake area A is a function of the water level h, A(h), which is determined by the water level-area hydrological relationship curve (i.e., the hyposometric curve you mentioned). This curve is shown in Appendix Figure A2.

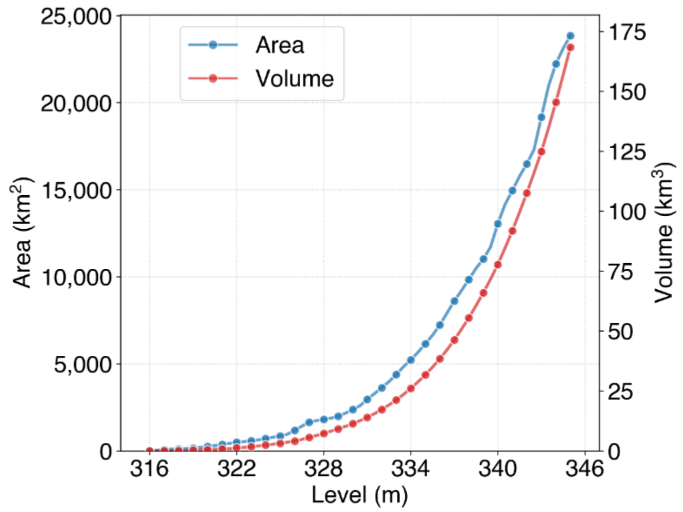

Figure A2. Stage-area and stage-volume relationships for Lake Balkhash. The blue line represents the relationship between water level and surface area (left axis), while the red line indicates the relationship between water level and storage volume (right axis). Data derived from Myrzakhmetov et al. (2022).

(12) Line 194: I understand including groundwater component is challenging. However, groundwater table decline is also a major contributor to water scarcity in arid regions. Therefore, evidence supporting the claim that groundwater has a relatively minor contribution in this study area should be provided here (e.g., relevant observations, previous studies, or sensitivity analyses).

Response: You raised a very important question. We have added references (Deng et al., 2011; Wang et al., 2022) to the section "2.3.3 Lake System…" to support our hypothesis that groundwater exchange accounts for a small proportion of the overall lake water balance in this study area.

(13) Figure 5: The figure resolution needs to be improved. In addition, consider segmenting the time series into calibration and validation periods using shaded boxes. The figure caption should also be revised to "Comparison between observed and simulated streamflow." Note that runoff in an open channel should be referred to as streamflow, not runoff.

Response: We have updated the image to a higher resolution. We have also adopted your suggestion to use a gray shaded background to distinguish between the calibration and validation periods, as explained in the figure captions. Furthermore, we have uniformly changed "runoff" in the figure to the more accurate "streamflow".

(14) Line 341: Which satellite altimetry & optical data was used to validate the reconstructed water storage? This needs to be clearly stated in section 2.2.

Response: We have explicitly added the source of the water level data used for verification in section "2.2 Datasets": "Historical gauge observations from 1931 to 2015 were obtained from Duan et al. (2020), while recent data (2016-2024) were extended using satellite altimetry products from the Global Reservoirs and Lakes Monitor (G-REALM)."

We would like to express our sincerest gratitude once again for your valuable time in reviewing our manuscript and providing such insightful feedback. We believe that the quality of the manuscript has been significantly improved under your guidance.

References

Deng, M., Wang, Z., and Wang, J.: Analysis of Balkhash Lake ecological water level evolvement and its regulation strategy, Journal of Hydraulic Engineering, 42, 403–413, 2011.

Duan, W., Zou, S., Chen, Y., Nover, D., Fang, G., and Wang, Y.: Sustainable water management for cross-border resources: The balkhash lake basin of central asia, 1931–2015, J Cleaner Prod, 263, 121614, https://doi.org/10.1016/j.jclepro.2020.121614, 2020.

Forgrave, R., Evenson, G. R., Golden, H. E., Christensen, J. R., Lane, C. R., Wu, Q., D'Amico, E., and Prenger, J.: Wetland-mediated nitrate reductions attenuate downstream: Insights from a modeling study, Journal of Environmental Management, 370, 122500, https://doi.org/10.1016/j.jenvman.2024.122500, 2024.

Harris, I.: CRU JRA v3.0: A forcings dataset of gridded land surface blend of Climatic Research Unit (CRU) and Japanese reanalysis (JRA) data; 1901 - 2024., 2025.

Ho, C.-C., Shih, P.-C., Chiang, L.-C., Lin, Y.-X., Liu, H.-F., Chang, W.-G., and Lin, G.-Z.: Mitigating non-point source pollution in tea plantations: SWAT modeling and field validation of slow-release fertilizers, Environ Monit Assess, 197, 810, https://doi.org/10.1007/s10661-025-14217-w, 2025.

Li, L., Long, D., Wang, Y., and Woolway, R. I.: Global dominance of seasonality in shaping lake-surface-extent dynamics, Nature, 642, 361–368, https://doi.org/10.1038/s41586-025-09046-3, 2025.

Myrzakhmetov, A., Dostay, Z., Alimkulov, S., Tursunova, A., and Sarsenova, I.: Level regime of balkhash lake as the indicator of the state of the environmental ecosystems of the region, Paddy Water Environ, 20, 315–323, https://doi.org/10.1007/s10333-022-00890-x, 2022.

Sánchez-Gómez, A., Bieger, K., Schürz, C., Rodríguez-Castellanos, J. M., and Molina-Navarro, E.: Multi-spatial and multi-criteria calibration to guarantee a robust SWAT+ hydrological model in a large and heterogeneous catchment, CATENA, 261, 109508, https://doi.org/10.1016/j.catena.2025.109508, 2025.

Wang, Z., Huang, Y., Liu, T., Zhong, R., Zan, C., and Wang, X.: Analyzing the water balance of Lake Balkhash and its influencing factors, Arid Zone Research, 39, 400–409, 2022.