Large-scale atmospheric drivers of recent hydroclimatic changes in the Horn of Africa: evidence from the Awash Basin terminal lake system, Ethiopia

Juan P. Sierra[1,2], Pierre Brigode[2], Pape Saara Ngom[2], Pascale Braconnot[3], Christian Le Carlier De Veslud[2], Camille Bouchez[2], Sly Wongchuig[4], Mohamed Osman Awaleh[5], Mohamed Jalludin[5], Natacha Volto[6], Eric Chaumillon[6], and Marie Revel[1]

[1]Géoazur, Université Côte d'Azur, CNRS, OCA, IRD, Valbonne, 06152, France
[2]Univ Rennes, CNRS, Géosciences Rennes, UMR 6118, 35000 Rennes, France
[3]Laboratoire des Sciences du Climat et de l'Environnement, Institut Pierre-Simon Laplace UMR CEA, CNRS, UVSQ, Université Paris-Saclay, Gif-sur-Yvette, 91190, France
[4]Université de Toulouse, LEGOS, CNES/CNRS/IRD/UT, Toulouse, France
[5]Institut des Sciences de la Terre, Centre d'Etudes et de Recherches de Djibouti (CERD), Route de l'aéroport B.P. 486, Djibouti, ville, Republic of Djibouti.
[6]La Rochelle Université, CNRS, UMR 7266 LIENSs, La Rochelle, France

Correspondence: Juan P. Sierra (juanpablosierrap@gmail.com)

Response to the reviewer 2

General comments

The manuscript "Large-scale atmospheric drivers of recent hydroclimatic changes in the Horn of Africa: evidence from the Awash Basin terminal lake system, Ethiopia" by Sierra et al., addresses a timely and relevant topic concerning hydroclimatic variability and lake-level changes in the Awash Basin. One of its main strengths is the combination of multiple datasets, methodological approaches, and perspectives, which allows the system to be examined from several complementary angles. I found this broad approach potentially valuable and think the study could provide useful insights into the climatic and hydrological controls on recent changes in the basin.

At present, however, I have several substantial concerns regarding both the methodological basis of the analysis and the way the results are presented. Most importantly, the hydrological model plays a central role in the paper and underpins several of its main conclusions, yet its formulation, calibration, validation, parameter selection, sensitivity, and performance are described only very briefly. This makes it difficult to assess how much confidence can be placed in the model results, particularly given the apparent change in model performance around the large lake-level rise in 2020. Related assumptions concerning evaporation, groundwater and transmission losses, precipitation biases, and the treatment of the atmospheric water budget also require stronger justification and, where possible, explicit evaluation of their associated uncertainties and residuals.

More generally, the manuscript currently contains a very large number of datasets, analyses, and relatively short results subsections. While this diversity is one of the study's strengths, it also makes the main scientific narrative difficult to follow. In several places, results are presented only briefly before the text moves directly into interpretation and comparison with previous

studies, making it difficult to distinguish clearly between what the analyses show and how the authors interpret those findings. This is particularly noticeable in the later parts of Sect. 3 and across Figs. 8–14. I therefore encourage the authors to revise the presentation: more clearly distinguish results from discussion even if the two remain within a combined section, and consider moving some of the supporting analyses and figures to the Supplementary Material. Doing so would help bring the central findings and the evidence supporting them much more clearly to the foreground.

Overall, I think the manuscript makes a very interesting contribution, but substantial revision is needed before the robustness of its main conclusions can be properly assessed.

Response

We thank the reviewer for the careful reading of our manuscript and for the constructive and detailed comments. We are grateful for the recognition that the multi-dataset, multi-method approach is a strength of the study, and we agree that the manuscript can be substantially improved in both its methodological transparency and its presentation. Below we respond to the general comments; specific comments are addressed point by point in the following section.

On the description of the hydrological model. We fully agree with the reviewer that the description of the hydrological model is too brief in the present version. In the revised manuscript we expand this considerably, providing: (i) a general description of the model formulation, including the definition and physical interpretation of each calibrated parameter; (ii) the calibration and validation protocol, with the periods used, the objective function, and the resulting parameter values; (iii) a sensitivity analysis of the calibrated parameters; and (iv) performance metrics for calibration and validation sub-periods. We agree that the change in model performance around 2020 deserves direct attention rather than being passed over, and in the new version of the document we present and discuss it openly.

On the role of the hydrological model in supporting our conclusions. Here we would like to offer a clarification, and we take the reviewer's reading as a clear indication that the current text does not convey the structure of our argument well enough. The hydrological model does not underpin the main conclusions of the study. Its role is deliberately limited: it is used to extend the record of lake area backwards in time before 1998, a period for which observational coverage in the region is sparse. This longer record allows us to place the recent trends in a longer context and to corroborate that the basin was in a dry phase with low lake levels before 1995. It is, in other words, a contextual and corroborative element rather than a load-bearing one.

The central conclusions of the manuscript—that the Awash Basin has become wetter during the July–September season, that this reflects changes in both regional and large-scale atmospheric circulation and in the surface hydrology of the basin, and that these changes have contributed to the expansion of the Afambo–Gemeri–Abhe lake system in recent decades—are supported by several independent lines of evidence that converge on the same result. These include (i) time-series analyses of lake-surface areas derived from satellite imagery; (ii) analyses of surface hydrology based on different precipitation datasets and evapotranspiration products; and (iii) trend analyses using precipitation products and the ERA5 atmospheric reanalysis, including variables such as moisture transport. We have also examined trends in the Normalized Vegetation Index using Landsat data (not shown), revealing a trend toward increased greenness in the basin that is consistent with positive rainfall trends. These

lines of evidence are independent of the hydrological model, and the main conclusions would remain unchanged if the model were removed entirely.

That said, we recognise that this hierarchy of evidence is not made sufficiently explicit in the current manuscript. In the revised version we will state clearly, both in the Methods and at the start of the results, which findings derive from observations and which from the model, and we will qualify the model-based inferences accordingly so that the reader can weigh them appropriately.

On the structure and presentation. We agree entirely with this comment. In the revised manuscript we will restructure the later part of Sect. 3 to separate more clearly what the analyses show from how we interpret them, synthesise the atmospheric analyses into a shorter and more consolidated subsections, and move several of the supporting analyses and figures (currently among Figs. 8–14) to the Supplementary Material. We believe this will bring the central findings and their supporting evidence more clearly to the foreground, as the reviewer recommends.

We address each specific comment individually below.

Major comments

Reviewer comment: *"2.2.2: This section lacks details, and assumes both that readers are familiar with the model, and that Figure 2 is enough to trust the model. I disagree with these two: Can you elaborate on the model? What are the two parameters? How were they chosen? Was the model calibrated (besides the factor 2 for the high flows)? Validated? How was the factor 2 for the high flows decided? Were there any sensitivity tests for the model parameters and for the correction of the high flows? How were the other precipitation datasets used in the model? What are the metrics used for the calibration and validation and how do you define a success of the model? What are the calibration objectives (match total volumes, high flows, specific peaks, etc.)? How is the model performance over the gauges for which it was not calibrated? Does this matter for your study?"*

Response

We thank the reviewer for this important comment. We agree that the description of the GR2M model in the original manuscript was too brief and that Fig. 2 alone does not provide sufficient information for the reader regarding the model calibration/validation process. We have therefore expanded the description of the hydrological model and its validation/calibration strategy in the main text.

We now describe the conceptual structure of GR2M, including its two reservoirs and the role of its two calibrated parameters. We also clarify the input data used to drive the model and, importantly, explain how the two parameters were determined for the Awash Basin. The revised description also distinguishes the calibrated parameters from the fixed components of the standard GR2M formulation.

Reviewer comment: *"How was the factor 2 for the high flows decided? Were there any sensitivity tests for the model parameters and for the correction of the high flows?"*

Response

We thank the reviewer for raising this point. We agree that the procedure used to select the high-flow correction factor should be described more clearly. We also identified an error in the original manuscript: the factor was reported as 2, whereas the factor actually used in the analysis was 1.8.

Because the available observed streamflow record at Tendaho is limited, we did not use the high-flow correction factor as an additional calibrated parameter of GR2M. Instead, we treated it as an empirical correction to account for the systematic underestimation of the upper tail of simulated streamflow. As shown in the new supplementary (here Fig. 1), GR2M reproduces the general range and temporal variability of observed Tendaho discharge, but its performance deteriorates for the most extreme flows, particularly for values above the 90th percentile of the observed record. We therefore tested the sensitivity of the subsequent lake simulations to different high-flow scaling factors.

Specifically, we tested scaling factors ranging from 1.0 to 2.0. For each factor, GR2M discharge values exceeding the 90th percentile were multiplied by the corresponding factor, while lower flows were left unchanged. The resulting discharge series was then used to force the lake water-balance model. Since our final objective is to reproduce the observed evolution of the Afambo, Gemeri and Abhe lake surfaces, and given the limited length of the observed Tendaho discharge record, we evaluated the effect of each correction factor using the lake observations rather than relying solely on the limited streamflow record.

For each scaling factor, we calculated the root mean squared error (RMSE) between the simulated and observed lake-surface areas over the complete available time series for the three lakes, considering the combined lake-surface series. The sensitivity analysis showed that a scaling factor of 1.8 provided the lowest overall RMSE, and this is therefore the factor adopted in the final simulations. The corresponding simulations obtained with the different scaling factors are now shown in the Supplementary Material, allowing the effect of the high-flow correction on the three lake-surface time series to be directly assessed.

We have revised the manuscript to correct the reported value from 2 to 1.8 and to clarify both the procedure used to select this factor and its role in the modelling framework. We have also added Fig. 1 to supplementary material.

Reviewer comment: *"How were the other precipitation datasets used in the model?"*

Response

We used CHIRPSv2 precipitation as the sole precipitation input to the GR2M hydrological model. Precipitation was spatially aggregated over the Awash Basin using the basin boundary as a spatial mask. This choice was based on previous comparisons between precipitation from several gridded datasets and available in-situ rainfall observations across the Awash Basin (Fig. 2). These comparisons indicated that CHIRPS provided the most consistent representation of observed precipitation over the basin.

We also performed additional tests using MSWEP and MSWXP as precipitation inputs to GR2M. In these experiments, the resulting model performance was lower than that obtained with CHIRPS, further supporting our choice of CHIRPS for the hydrological reconstruction.

The other precipitation products were therefore not used as alternative inputs to the hydrological model. Instead, we used an ensemble of precipitation products, including CHIRPSv2, CMORPH, TRMM3B42, and CenTrends to assess the robustness and spread of the estimated precipitation trends and interannual variability over the Awash Basin. This multi-product approach allows us to distinguish signals that are consistent across datasets from those that may depend on a particular precipitation product, while keeping the hydrological reconstruction based on the precipitation dataset that performed best against the available observations.

Reviewer comment: *"What are the metrics used for the calibration and validation and how do you define a success of the model? What are the calibration objectives (match total volumes, high flows, specific peaks, etc.)? How is the model performance over the gauges for which it was not calibrated? Does this matter for your study?"*

Response

We thank the reviewer for raising these important points. We have clarified the calibration and validation procedure, the associated performance criteria, and the scope of the model evaluation in the revised manuscript.

The GR2M model was calibrated against observed monthly discharge at Tendaho using the Michel (1991) calibration algorithm implemented in the airGR package (Coron et al., 2017), following a four-year warm-up period. The objective function was the Kling–Gupta efficiency (KGE; Gupta et al. (2009)), which combines correlation, bias in mean discharge, and differences in discharge variability between simulated and observed time series. The optimization therefore aimed to reproduce the overall monthly discharge dynamics, rather than explicitly targeting particular flood peaks or cumulative discharge volumes. The ideal KGE value is 1, but there is no universal threshold defining successful model performance; we assess performance in relation to the intended application and the results obtained during validation.

The GR2M model was calibrated over the full 1990–2004 period, and the resulting parameter set was used for the final discharge simulation. To assess the sensitivity of the calibrated parameters to the choice of calibration period, we conducted two additional split-period experiments (Table 1). In the first experiment, calibration was performed over 1990–1998 and evaluation over 1999–2004; in the second, these periods were reversed. The resulting parameter estimates and KGE scores were broadly comparable in terms of model performance, although the differences in (X_1) indicate some sensitivity of parameter estimates to the selected calibration period. The KGE obtained from calibration over the full period was within the range of the calibration scores from the split-period experiments. These tests provide an assessment of temporal robustness, although they do not constitute an independent validation of the final simulation. We have added this table to the supplementary material.

Regarding performance at other gauges, the model was configured as a lumped model for the catchment draining to Tendaho, which provides the discharge boundary condition for our lake-system analysis. We did not independently calibrate or validate this configuration at other gauging stations, as these would require separate delineations of their contributing catchments and corresponding model applications. Consequently, our evaluation does not establish the spatial transferability of the calibrated parameters across the Awash Basin. However, our primary objective is to reconstruct the long-term inflow to

the Afambo–Gemeri–Abhe lake system, for which the Tendaho record provides the relevant observational constraint. We acknowledge this limitation and have clarified the scope of the model evaluation accordingly.

Table 1. Calibration and validation performance of the hydrological model GR2M under two split-period experiments and calibration over the full period (1990–2004).

Experiment	Calibration period	Validation period	X_1 (mm)	X_2	KGE (calibration)	KGE (validation)
Split 1	1990–1998	1999–2004	1743	0.713	0.738	0.666
Split 2	1999–2004	1990–1998	2836	0.702	0.722	0.677
Full period	1990–2004	—	2039	0.712	0.724	—

Reviewer comment: *"L177: What do you mean by "bias-correction"? How come there is still a positive bias after it?"*

Response

We agree that the term "bias correction" was misleading. The correction applied to the GR2M simulations was not intended to remove the overall bias of the model. Rather, it was a high-flow adjustment designed to improve the representation of the upper tail of the simulated discharge distribution, which was systematically underestimated by GR2M. Because this adjustment is applied only to the upper tail, it does not necessarily eliminate the overall mean bias. In fact, the remaining positive bias (+11.59%) mainly reflects the increased magnitude of wet-season flows during anomalously wet years after the high-flow adjustment. We have therefore replaced "bias correction" with "high-flow adjustment" throughout the manuscript and clarified its purpose and implications.

Reviewer comment: *"L192–197: Can you think of a reason why GLEAM should perform OK for the other lakes but not for Lake Abhe? Are there other PET estimates for the region in contrast with GLEAM? In that case, why should we trust it in other places? Can the model mismatch come from a different factor rather than evaporation only?"*

Response

We agree that the original text was not sufficiently clear about the selection of GLEAM or about the interpretation of the evaporation adjustment.

First, GLEAM was not selected because we assumed that it performs better specifically over the three lakes. We evaluated several available evapotranspiration/evaporation products, including WatGAP3, MOD16A2, PCR-GLOBWB, HTESSEL, and LISFLOOD (we include all these datasets in the basin-scale analysis in original's manuscript Figure 4). We selected GLEAM for three main reasons. (i) GLEAM estimates evaporation and evapotranspiration using a physically based framework including the Penman–Monteith approach, which is appropriate for representing atmospheric evaporative demand. (ii) Chen et al. (2024) evaluated combinations of precipitation and evapotranspiration datasets for closing the water balance over Ethiopia and found that the combination including GLEAM evapotranspiration provided the best-performing dataset combination for

estimating surface runoff. Their analysis also reported a relatively small water-balance closure error of 46.5 $\mathrm{mm\,y^{-1}}$ (5.6%) at the national scale. (iii) A comparison of GR2M discharge simulations at Tendaho using different precipitation and potential evapotranspiration datasets showed that model performance was more sensitive to the choice of precipitation dataset than to that of potential evapotranspiration. Among the tested combinations, using GLEAM potential evapotranspiration together with CHIRPS v2 precipitation yielded one of the best model performances (not shown). These results provided additional support for using GLEAM as the evapotranspiration forcing in our study.

Second, the factor of 1.5 should not be interpreted as evidence that GLEAM systematically underestimates evaporation specifically over Lake Abhe. The value for Lake Abhe was used as an example to illustrate the magnitude of the adjustment. In our lake-balance model, the same multiplicative adjustment was applied consistently to all three lakes. The factor was selected through sensitivity analysis (original manuscript Fig. S4), rather than derived from a direct comparison between GLEAM and an independent evaporation measurement at Lake Abhe.

We also agree that the discrepancy between simulated and observed lake areas cannot necessarily be attributed to evaporation alone. Other sources of uncertainty include simulated Awash River discharge, water inflows from the Gobaad region (which are not taken into account in our lumped model), uncertainties in lake storage–area relationships and bathymetry, and anthropogenic modifications of the downstream flow regime (as shown below in the following comments). We therefore interpret the evaporation adjustment as a sensitivity-based treatment of an important but uncertain component of the lake water balance, rather than as a definitive correction of a GLEAM bias.

We have revised the manuscript to clarify both the rationale for selecting GLEAM and the fact that the 1.5 factor represents an empirical adjustment applied to all three lakes rather than a Lake Abhe-specific correction.

Reviewer comment: *"Groundwater: This is mentioned as a variable in Eq. 1 but later said to be neglected. Can you reinforce the fact it is negligible? How about the sandy river bottoms in the lower basin next to Lake Abhe? This factor can be super important in arid, sandy regions (Knighton and Nanson, 1994; Morin et al., 2009)."*

Response

We thank the reviewer for raising this important point. We agree that our original wording was too categorical and did not sufficiently distinguish between regional groundwater discharge to the river and localized river-to-groundwater exchange through alluvial sediments.

Our assumption is supported by regional hydrogeological evidence concerning groundwater contributions to the Awash River. Kebede et al. (2021), using environmental tracers and piezometric observations, found no detectable groundwater inflow to the Awash between Tendaho and Lake Abhe and showed that the middle and lower Awash is predominantly a losing river. This indicates that regional groundwater discharge is unlikely to represent a major source term for the river–lake water balance in our study area. The hydrogeochemical and isotopic study of the Lake Abhe geothermal field by Awaleh et al. (2015) further demonstrates the presence of distinct groundwater and geothermal systems in the Lake Abhe area, including hot springs, wells and boreholes. Their results distinguish the deeper, older geothermal waters from surface waters and local aquifers recharged

by modern precipitation, indicating that the groundwater systems in the Abhe area are hydrogeologically complex. However, the study does not indicate a substantial direct groundwater contribution from these geothermal systems to the Awash River or the connected Afambo–Gemeri–Abhe lake system.

We nevertheless agree with the reviewer that this does not imply an absence of groundwater–surface-water exchange at the local scale or during transient hydrological events, such as floods, when changes in river stage may temporarily modify the direction and magnitude of exchange through alluvial sediments. In particular, the lower Awash contains alluvial and sandy river sediments, where river-to-groundwater transmission losses are expected under arid conditions. This process is well documented in other arid and hyperarid river systems. As you mention, Morin et al. (2009) explicitly demonstrated that infiltration through the channel bed of the Kuiseb River can produce transmission losses and recharge shallow alluvial aquifers. Knighton and Nanson (1994) likewise documented the importance of transmission losses in arid river systems. We have added these references to the discussion.

The key distinction is therefore between the direction and scale of the exchange: available evidence indicates that substantial regional groundwater inflow to the Awash River is limited, whereas localized river-to-groundwater losses through alluvial channel beds can occur. We cannot quantify the latter reliably for the lower Awash or near Lake Abhe because the available observations do not provide sufficiently resolved discharge measurements, groundwater levels or river–aquifer hydraulic gradients along the relevant reaches.

We have revised the manuscript accordingly. Rather than stating that groundwater fluxes are simply "negligible," we now explain that they are omitted as a first-order simplification of the regional water balance, while explicitly acknowledging localized river-to-groundwater transmission losses as an additional source of uncertainty.

Reviewer comment: *"3.1: There seems to be a very big discrepancy between the model performance before the big 2020 rise and after it. This odd behaviour needs to be clarified by both describing the model validation process, and finding the reasons for this discrepancy."*

Response

We thank the reviewer for highlighting this discrepancy. We agree that the change in model performance around 2020 warrants further investigation. We have therefore expanded Section 3.1 to reconstruct the evolution of the three lakes over the past 40 years and investigate potential links with precipitation variability and water management in the lower Awash Basin.

Our initial hypothesis was that GR2M might fail to reproduce the effects of particularly intense precipitation events occurring during the last 5 years. However, daily CHIRPS precipitation data did not provide sufficient evidence to support this explanation. We therefore examined Landsat-derived surface areas of Tendaho Reservoir and the downstream lakes Afambo–Gemeri and Abhe (Fig. 3). Between 2013 and 2024, we identified 14 events during which the reservoir surface area exceeded its approximate maximum-capacity reference of 170 km^2 (Woldeyohanis, 2010) before subsequently decreasing. Single-month exceedances were excluded because some appeared to be artifacts associated with missing satellite observations, while their

short duration made downstream responses difficult to assess. Although surface-area changes cannot directly confirm reservoir releases, these events provide an indication of potential changes in reservoir storage and operation.

Before 2016, the three lakes generally exhibited coherent variations. From 2016 onward, their behavior increasingly diverged: Afambo–Gemeri often responded to Tendaho exceedances after a delay of several months, while Lake Abhe continued to decrease. This pattern was particularly evident in late 2016 and during 2017–2018 (Fig. 3a). Following the exceptional expansion of all three water bodies in late 2019 and 2020, the downstream response appeared to change further, with increases in Afambo–Gemeri followed by a more pronounced response in Lake Abhe. Meanwhile, Tendaho exceedance events became both longer and more frequent after 2020. Their mean duration increased from 3.5 ± 0.9 months before 2020 to 4.85 ± 2.2 months after 2020, while the mean interval between events decreased from 9.5 ± 1.5 months to 6.0 ± 2.8 months, respectively (mean $\pm$ standard deviation; Fig. 3b).

To further investigate the downstream response, we performed a composite analysis of the 14 exceedance events, examining surface-area changes from three months before to ten months after each event (Fig. 3c). Surface areas were normalized by their respective three-month pre-event means. The composite shows that decreases in Tendaho Reservoir surface area are generally accompanied by increases in Afambo–Gemeri and, subsequently, Lake Abhe, consistent with a progressive downstream response. These observations suggest that reservoir regulation and associated water transfers may have contributed to changes in downstream lake dynamics, particularly after 2020.

Taken together, these results suggest that the post-2020 discrepancy may partly reflect changes in reservoir operation rather than solely limitations in the rainfall–runoff model. GR2M does not explicitly represent reservoir operations, controlled releases, or irrigation withdrawals; consequently, these processes could contribute to the divergence between simulated and observed lake dynamics. To investigate this hypothesis further, we conducted an additional sensitivity experiment to assess whether accounting for potential excess water from Tendaho Reservoir could improve the simulation of downstream lake dynamics.

Because the SRTM topographic data were acquired in 2000, eight years before the construction of Tendaho Dam, we used the pre-dam topography of the reservoir area as a proxy for its bathymetry (Fig. 4a). From these data, we derived a hypsometric curve relating water surface area to volume (Fig. 4b), which allowed us to estimate the excess water volume associated with reservoir surface areas exceeding the approximate maximum-capacity threshold. We then added these estimated excess volumes to the GR2M-simulated Awash River discharge as a simplified proxy for water potentially released or spilled downstream when the reservoir exceeds its estimated maximum capacity. The resulting discharge was then used as input to our mass-balance-based lake model.

This simplified experiment reproduced the main features of the observed lake dynamics after 2020, particularly the abrupt expansion of all three lakes in 2020 and 2024 (Fig. 4c). However, the experiment has important limitations: the estimated excess volumes are introduced as additional water without explicitly accounting for the corresponding reservoir water balance, and the entire volume is added to the river within a single monthly time step rather than being distributed over time according to actual release operations. The experiment should therefore not be interpreted as a physically consistent reconstruction of reservoir releases. Nevertheless, it demonstrates that introducing additional water volumes at the appropriate times can reproduce the

major post-2020 lake expansions, supporting the hypothesis that reservoir management may have contributed to the observed changes. Direct confirmation of this mechanism would require independent information on reservoir operations and water releases.

In accordance with the previous discussion, we have amended Section 3.1 as follows. First of all, we quantified the lake model's performance over the 2018–2024 period to illustrate the deterioration in model performance during recent years. We also added a brief discussion exploring the possible influence of Tendaho Dam on this deterioration in a new appendix, supported by Figures 3 and 4.

Reviewer comment: *"One of the strengths of this manuscript is the combination of multiple datasets, methodological approaches, and perspectives, which allows the problem to be examined from several complementary angles. At the same time, however, the large number of datasets, methods, and relatively short results subsections makes the overall flow of the paper difficult to follow. In particular, Figs. 8–12 are covered within only a few sentences, and it is often unclear where the presentation of results ends and the discussion begins. Please make a clearer distinction between the two, even if Results and Discussion remain combined. The reader should be able to distinguish clearly between what was found and the authors' interpretation of what these findings mean, their possible causes, and how they relate to previous studies.*

I also suggest considering whether all of Figs. 8–14 are needed in the main text, particularly as many of their main messages are ultimately synthesised in Fig. 14. Some of these figures could perhaps be moved to the Supplementary Material/Appendix, while relevant background material from the opening paragraphs of these subsections could instead be incorporated into the Introduction. This is particularly relevant for Sect. 3.3, where the results themselves are described only very briefly before the text moves into an extended discussion. It may therefore be clearer to treat much of this section explicitly as Discussion, while summarising the key results more concisely and presenting only the main findings in one or two figures."

Response

We thank the reviewer for this constructive comment. We agree that the combination of multiple datasets and methodological approaches is an important strength of the study, but that the presentation of these results can make the overall narrative difficult to follow, particularly in Sect. 3.3 and Figs. 8–14.

In the revised manuscript, we have restructured these sections to provide a clearer distinction between the results and their interpretation, while retaining the combined Results and Discussion format. We first present the main findings more explicitly and concisely, focusing on what is directly shown by the different datasets and analyses, and then discuss their possible causes, physical interpretation, and consistency with previous studies. We have also revised the transitions between subsections to make the progression from observations to interpretation more explicit.

Following the reviewer's suggestion, we have also reassessed the figures currently presented in the main text, particularly Figs. 8–14. We will remove or move to the Supplementary Material figures that provide complementary information or repeat results whose main message is already captured elsewhere. The main text will retain only the figures that are essential for

supporting the central narrative and the key atmospheric mechanisms identified in the study. Where appropriate, background information currently presented at the beginning of these subsections will instead be incorporated into the Introduction.

Overall, these revisions will reduce redundancy, improve the balance between results and discussion, and make the main line of evidence—from the observed hydroclimatic changes to their potential large-scale atmospheric drivers—easier to follow.

Specific comments

Reviewer comment: *L26: "have" should be "may have / may impose" or be supported by a reference.*

Response

We agree with the reviewer and we have changed this phrase according to the suggestion.

Reviewer comment: *"L32: What do you mean by "recurrent droughts"? That both the short and long rains are low, or that there are two consecutive drought years?"*

Response

Thank you for pointing this out. By "recurrent droughts," we referred to drought episodes that recur at multi-year intervals, rather than to consecutive drought years or the simultaneous failure of different rainy seasons. Wolde-Georgis (1997), based on previous analyses of rainfall deviations, reports recurrence intervals of 3–5 and 6–8 years in northern Ethiopia and approximately 8–10 years for Ethiopia as a whole. We have therefore clarified the wording in the manuscript and revised the sentence to: "Drought episodes have historically recurred at intervals of approximately 8–10 years across Ethiopia (Abate, 1994; Wolde-Georgis, 1997) . . . "

Reviewer comment: *"L33: 80% of the annual mean – is this for the entire country or the inner parts? These numbers are in contrast to those shown by Palmer et al. (2023). I think the clarification is done in L64, where the same numbers are stated for only the Awash basin. Please reconcile these two."*

Response

Thank you for pointing this out. We agree that the original wording could be interpreted as suggesting that nearly 80% of annual rainfall falls during June–September throughout the entire country. The literature indicates substantial spatial variability in rainfall seasonality across Ethiopia, with the June–September rainy season accounting for approximately 50–80% of annual rainfall across much of the country (Billi et al., 2015; Nicholson, 2017). We have therefore replaced "nearly 80%" with "approximately 50–80%" and clarified that this range applies across much of Ethiopia rather than uniformly to the entire country.

We also checked Palmer et al. (2023), which emphasizes the strong spatial variation in the number and timing of rainy seasons across Ethiopia and the Horn of Africa, rather than providing a single national percentage of annual rainfall occurring

during June–September. We therefore do not consider the two statements contradictory, but agree that our original wording was insufficiently precise. We have revised the text accordingly.

Reviewer comment: *"L97: One of the things that Dinku et al. (2018) mention is the bias CHIRPS data have over dry (and in some cases mountainous) regions. Moreover, Dinku et al. (2011) specifically describe the problem of below cloud rainfall evaporation in dry regions. Finally, GLEAM uses MSWEP precipitation as input. This means that by using other precipitation datasets you also inherit some error in the ET coming from the precipitation product used in GLEAM. Please address these issues."*

Response

Thank you for raising this important point. We agree that our original statement was too broad and did not sufficiently acknowledge the spatially varying uncertainties of satellite precipitation estimates over Ethiopia.

Dinku et al. (2018) show that CHIRPS performs well overall over Ethiopia, with high skill at both decadal and monthly timescales, but also report substantial spatial variability in performance. In particular, they identify lower skill over southern Ethiopia and challenges associated with mountainous regions. Dinku et al. (2018) further showed that satellite rainfall products can underestimate rainfall over the Ethiopian highlands, partly because of difficulties in detecting warm orographic rainfall, while overestimation can occur over dry regions, potentially owing to sub-cloud evaporation. We have therefore revised the text to avoid implying that CHIRPS is uniformly reliable across Ethiopia and now explicitly acknowledge these limitations.

We also clarify that CHIRPS was selected as the primary precipitation dataset after comparison with additional precipitation products and gauge observations, while the spread among precipitation products was used to characterize uncertainty in the precipitation trends and variability.

Regarding the use of GLEAM, we agree that this introduces an additional source of dependence that should be acknowledged. GLEAM v4.2a uses MSWEP precipitation as an input, meaning that uncertainties in the precipitation forcing can propagate into its evapotranspiration estimates. We have added a statement in the revised manuscript noting that comparisons between GLEAM evapotranspiration and precipitation products are therefore not fully independent.

The corresponding text has been revised accordingly.

Reviewer comment: *"L99: "TRMM 3B42 V7" – do you mean IMERG V7?"*

Response

We confirm that the precipitation dataset used in our analysis is TRMM 3B41 V7, not GPM IMERG V7. This dataset covers the period from 1 January 1998 to 31 December 2019, with a temporal resolution of 3 hours and a horizontal resolution of 0.25°.

Reviewer comment: *"L118–120: Are there temperature discrepancies when using CRU data for mean T and ERA5 for T? Why do separate these into two datasets?"*

Response

Thank you for raising this point. We agree that the use of two different temperature datasets requires clarification. We used CRU TS v4.09 for monthly mean near-surface temperature because it is an observation-based gridded dataset derived from station observations and spatial interpolation, whereas ERA5 is a reanalysis that combines observations with a numerical weather model and data assimilation. The two datasets therefore provide complementary estimates of near-surface temperature based on different methodologies. CRU TS and ERA5 can be considered independent data products for the purpose of this analysis, although we acknowledge that some observational information may be shared between their underlying data sources.

To assess whether the choice of dataset affects our results, we repeated the temperature analysis using ERA5 near-surface air temperature. The resulting temporal variability over the Awash Basin is highly consistent with that obtained from CRU TS, although ERA5 shows systematically higher temperature magnitudes. Importantly, the main temporal features and trends identified from CRU TS are also reproduced using ERA5 (new Figure 4). We have therefore added this comparison to the revised manuscript and clarified that CRU TS is used for the observationally based characterization of near-surface temperature, while ERA5 is used primarily to characterize the three-dimensional atmospheric state and its dynamical and thermodynamical mechanisms.

Reviewer comment: *"Sect. 2.1.7: Please explain: (1) How the Landsat-based extension (2021–2025) works; (2) what's the reliability of the data from Zhao et al. (2022), which are based on GSWD, to the region; (3) what is the period of the HydroWeb span, and how reliable it is. If it does not, what did you do for the previous period? (4) what HydroWeb product was used?"*

Response

(1) Landsat-based extension (2021–2025): We clarify that the 2021–2025 extension does not represent an extension of the GSWD product itself. The GSWD record ends in 2021, so we used Landsat 8 and 9 imagery acquired after 2021 to continue monitoring surface-water extent in the study area. We inspected Landsat scenes covering the main water bodies considered in this study (Lakes Abhe, Gemeri and Afambo, and the Tendaho reservoir) and delineated the water extent from these images. This allowed us to extend the satellite-based lake surface-area observations through 2025. We have clarified this procedure in Section 2.1.7 of the revised manuscript.

(2) Reliability of the Zhao et al. (2022) reconstruction: We thank the reviewer for asking us to clarify the reliability of this dataset. Zhao et al. (2022) developed their reconstruction specifically to address missing observations in the monthly GSWD product, which are particularly important during the early Landsat period. Their method combines annual GSWD water classifications, which contain substantially fewer missing pixels, with the long-term monthly water-area climatology derived from the original GSWD monthly observations. For years in which the annual year-round and/or seasonal water areas cannot be directly determined because of insufficient Landsat coverage, the corresponding values are linearly interpolated from adjacent years.

The reconstruction was evaluated by Zhao et al. (2022) against cloud-free monthly observations from the Global Reservoir Surface Area Dataset for 6,715 reservoirs. They reported an overall R2 of 0.97, a median relative bias of 1.6%, and a median

relative RMSE of 9.2%. Importantly, the reported agreement varies with water-body size: the reconstruction performed particularly well for lakes and reservoirs larger than 100 km^2 (R2 > 0.996), whereas substantially larger errors were found for very small water bodies (<0.01 km2). This indicates that the reliability of the reconstruction is partly dependent on lake size and the availability of Landsat observations.

Lake Abhe, with a surface area of approximately 250 km^2, is larger than the 100 km^2 class for which Zhao et al. (2022) reported the strongest validation statistics. Nevertheless, its large size makes it substantially different from the smallest lakes for which the reconstruction performs less well, and the Zhao et al. (2022) validation provides general evidence that the approach can capture lake-surface-area dynamics at relatively large spatial scales. We also note that these validation statistics are global and are not specific to the Afar Depression.

We therefore use the Zhao et al. (2022) reconstruction as a complementary source of information for the early period, when direct Landsat observations over the region are sparse, rather than treating every reconstructed monthly value as a direct observation. In particular, we acknowledge that uncertainty is higher for individual years or months with limited Landsat coverage, for which part of the reconstruction relies on interpolation and the long-term seasonal climatology. These limitations and their implications for the interpretation of the reconstructed time series are now explicitly discussed in the revised manuscript.

(3) What is the period of the HydroWeb span, and how reliable is it? If it does not, what did you do for the previous period? (4) What HydroWeb product was used?: We thank the reviewer for pointing out that the original manuscript did not sufficiently describe the temporal coverage, product, and specific purpose of the HydroWeb data. We have clarified this information in the revised manuscript.

HydroWeb observations were available only for Lake Abhe among the study water bodies. We used the HydroWeb Lake Abhe water-level time series (HydroWeb station/product ID 1300000001548), which covers 13 September 2010 to 20 November 2020. The product provides water-surface elevation derived from satellite radar altimetry, with corrections for solid Earth and pole tides, ionospheric and wet/dry tropospheric delays, and altimeter biases, and elevations referenced to the EGM2008 geoid. The time series is produced within the HydroWeb/SOLS framework developed by LEGOS and computed by CLS on behalf of CNES and the Copernicus Global Land service, with subsequent improvements supported by ESA and the Copernicus Climate Change Service.

Importantly, HydroWeb water-level observations were not used to extend or fill the lake surface-area time series. Their purpose in this study was to provide independent water-surface-elevation information for the reconstruction of Lake Abhe bathymetry. We combined HydroWeb water levels with SRTM topography, satellite-derived lake surface areas and shoreline contours, and in situ water-depth measurements collected during the NILAFAR field campaign to reconstruct the lake bathymetry. From this reconstruction, we derived the area–water-level and area–volume relationships used in the lake analysis.

Consequently, the absence of HydroWeb observations before September 2010 does not require a reconstruction of the earlier water-level record. For the earlier period, lake surface-area variability was characterized from the Landsat-derived observations and the monthly reconstructions of Zhao et al. (2022) as described in the manuscript. HydroWeb was used only as a source of water-surface-elevation constraints during the period for which altimetry observations are available.

The reliability of HydroWeb depends on the accuracy of satellite-altimetry water levels products, which has been evaluated in numerous studies. For example, Crétaux et al. (2011) reported an RMS difference of 2.6 cm and a correlation of 0.99 between Jason-1-derived water levels and in-situ gauge measurements for Lake Victoria over September 2004–March 2006. Crétaux and Birkett (2006) also reported RMS errors of approximately 3–5 cm for the Great Lakes. However, these results concern large lakes and should not be directly extrapolated to Lake Abhe. Ricko et al. (2012) showed that accuracy varies among water bodies and satellite-altimetry products, with RMS differences of approximately 0.20–1.41 m reported for narrower reservoirs and smaller lakes. More recent assessments likewise emphasize the influence of lake geometry, satellite mission characteristics and processing methods on the uncertainty of inland-water level estimates.

Reviewer comment: *"Sect. 2.2.1: This section seems to resolve one of the problems solved by Zhao et al. (2022), which is one of the input sources. Isn't it? Please explain."*

Response

We thank the reviewer for pointing out this potential ambiguity. We agree that the original presentation did not sufficiently distinguish the different purposes of the satellite-based reconstructions and the lake model.

Both approaches address the problem of incomplete lake observations, but they do so differently and provide complementary information. The reconstruction of Zhao et al. (2022) is a satellite-based product that addresses missing observations in the Landsat record by combining annual GSWD water classifications with the long-term monthly water-area climatology derived from the GSWD monthly product. We use these reconstructions as complementary observational estimates, particularly during periods with sparse Landsat coverage.

In contrast, our lake model provides an independent hydrologically constrained reconstruction based on the lake water balance, using basin hydrological inputs and atmospheric forcing. It does not reconstruct missing satellite images or rely on the same GSWD-based information as the Zhao et al. (2022) product. This distinction is particularly important for the period before 1998, when Landsat observations over the Afar Depression are very sparse. The lake model allows us to reconstruct lake-area dynamics during this period from the underlying water balance and therefore provides an independent estimate of lake evolution that can be compared with the satellite-based reconstructions.

We have revised Sect. 2.2.1 to clarify that the corrected GSWD/Landsat observations and the Zhao et al. (2022) reconstructions are used as observational constraints and for evaluation of the lake model, whereas the lake model provides an independent reconstruction based on hydrological and atmospheric forcing. We believe this distinction also makes clearer why both approaches are retained in the analysis.

Reviewer comment: *"Appendix: Do a and b correspond to x and y in the A1–A5 eqs.? Please make this clearer. Also, these equations assume no transmission losses. Can this be justified? How about irrigation abstractions along the rivers?"*

Response

We thank the reviewer for pointing out these two aspects. We have clarified the relationship between the parameters used in Eqs. A1–A5: X and Y represent the relative partitioning of the post-abstraction flow, while the effective routing parameters used in the model are $a = X(1-p)$ and $b = Y(1-p)$. Thus, a and b account jointly for flow partitioning and the reduction represented by p.

The model does not explicitly represent transmission losses along the river channels. Instead, p represents the effective fraction of Tendaho discharge removed before reaching the downstream lake system, primarily associated with irrigation abstractions but potentially also encompassing other unrepresented losses. We have clarified this assumption in Appendix A. Given the limited availability of discharge observations and independent information on irrigation withdrawals and channel losses, these processes cannot be robustly separated in the present model. The estimated temporal evolution of p is therefore interpreted as an effective flow reduction rather than as a direct measurement of irrigation abstraction.

Reviewer comment: *"Fig. S4 and hypsometric curves: The figure is not clear enough – what are the Altimetry radar contours? Can you state their elevations? Elevation contours seem unphysical; consider either smoothing them, or using a different approach, e.g., the one from Armon et al. (2020). Can you validate the hypsometry using external data (e.g., ICESat-2)?"*

Response

We thank the reviewer for these helpful suggestions. We agree that the original figure and caption did not sufficiently distinguish between the topographic elevation contours, satellite-derived shorelines and altimetry observations. We have revised the figure caption and expanded the methodological description to clarify these distinctions and better explain the reconstruction and its limitations.

Regarding the first point, the blue lines in Fig. S4a represent lake shorelines delineated from Landsat imagery acquired on different dates, each associated with a lake water-surface elevation derived from available HydroWeb radar-altimetry observations. HydroWeb combines measurements from multiple satellite missions to provide lake water-level time series. The blue lines therefore represent satellite-derived shoreline positions associated with altimetry-based water levels, rather than elevation contours measured directly by radar altimetry. We have clarified this distinction in the revised caption and specified the corresponding water-surface elevations in metres above sea level.

The red lines represent SRTM-derived topographic elevation contours. Their apparent irregularity primarily reflects the complex, low-relief topography along the southern and southwestern shores of Lake Abhe, where small depressions and plateaus occur over elevation differences of approximately 1 m. The limitations of the SRTM digital elevation model and the spatial interpolation procedure may also contribute to these patterns. We have expanded the methodological description to explain these features and clarify that the small-scale irregularities should not be interpreted as accurately resolved bathymetric features. Rather than applying indiscriminate smoothing, which could remove relevant topographic features and alter the reconstructed area–volume relationship, we have retained the underlying elevation information while making its limitations more explicit.

We also examined the publicly available 3D-LAKES bathymetry dataset of Huang et al. (2025). However, its representation of Lake Abhe did not adequately resolve the lake interior or reproduce the morphology of its northeastern margin, where geological structures, potentially including faults, may influence the lake-floor geometry. We therefore did not consider it a suitable replacement for our reconstruction.

Following the reviewer's suggestion, we additionally examined ICESat-2 ATL03 geolocated photon data to assess the consistency of our reconstructed bathymetry with an external dataset. Identifying lake-bottom elevations in ATL03 is challenging because the observations may include water-surface and potential bottom returns, as well as atmospheric photons and noise (Fig. 5). We interpolated selected photon observations in three dimensions and extracted an elevation profile along the same transect as that used for our reconstructed bathymetry (Fig. 6). The profiles show broadly consistent spatial variations after applying an approximately 6 m vertical shift to the ICESat-2-derived profile, although differences remain near the northeastern margin. The origin of this offset remains uncertain and requires further investigation of the vertical reference systems, acquisition dates and identification of the photon returns. We therefore interpret this comparison as a preliminary assessment of the consistency of the bathymetric profiles, rather than an independent validation of their absolute elevations. We have clarified this limitation in the revised manuscript.

Regarding the alternative methodology suggested by the reviewer, we acknowledge that our approach shares important conceptual principles with the Water Occurrence and Laser Profile (WOLP) method proposed by Armon et al. (2020). Both approaches combine satellite-derived information on lake extent with elevation constraints to reconstruct a continuous lake-floor surface in the absence of comprehensive bathymetric surveys. However, they differ in how these observations are integrated. Armon et al. (2020) establish an empirical relationship between long-term water-occurrence frequency derived from Landsat imagery and ICESat-2 ATL03 elevation measurements, which is then applied across the lake. In contrast, our approach associates shoreline positions delineated from Landsat imagery with HydroWeb water-surface elevations and combines these constraints with surrounding SRTM topography and three in-situ water-depth measurements collected during the NILAFAR field campaign. These complementary data are integrated through spatial interpolation to reconstruct the lake-floor geometry.

We therefore consider the existing approach appropriate for the objectives of this study and do not consider implementation of the full WOLP method necessary. Our reconstruction incorporates direct in-situ depth constraints alongside satellite-derived shoreline and altimetry information, and it provides satisfactory estimates of lake surface areas and volumes within our lake water-balance model. The broadly consistent spatial variations observed in the ICESat-2 profile provide an additional, although preliminary, consistency check. Together, these results support the suitability of the reconstruction for the water-balance analysis, while the unresolved vertical offset and local differences are acknowledged as limitations. We have included a brief discussion comparing our approach with that of Armon et al. (2020) , as well as our lake-bottom elevation results with data obtained from ICESat-2.

Reviewer comment: *"Sect. 2.3.2: Why don't you use GLEAM to assess ET here? Is this consistent with your use in the hydrological model?"*

Response

GLEAM was in fact also used to estimate evapotranspiration in the atmospheric water-budget analysis at basin scale, but it was inadvertently omitted from the list of ET products in Sect. 2.3.2. We have corrected this omission in the revised manuscript.

Using GLEAM in both the atmospheric water-budget analysis and the hydrological modelling is consistent. GLEAM was selected as the evapotranspiration forcing for the hydrological model based on the existing literature and previous assessments of ET products over Ethiopia. In the atmospheric water-budget analysis, we used GLEAM together with several other ET products to account for uncertainties associated with the choice of ET dataset. Thus, GLEAM provides continuity with the ET forcing used in the hydrological model, while the multi-product approach allows us to assess the robustness of the atmospheric moisture-budget results at basin scale to ET uncertainty.

The revised text now explicitly includes GLEAM among the ET products used in the atmospheric water-budget analysis.

Reviewer comment: *"Why not using surface pressure to top of the model (or if it's easier, 200 hPa)? I'm asking because in large parts of the region the surface is well above the 1000 hPa line. Do you take the lowermost layer or what? Additionally, given that the datasets are not equal, I would not expect a water balance closure. Can you quantify the residual using the different datasets?"*

Response

Thank you for pointing this out. We agree that using a fixed lower boundary of 1000 hPa is not appropriate over the elevated parts of the Awash Basin, because several of the lower pressure levels can lie below the local surface. In the original calculation, these below-ground pressure levels were inadvertently included in the vertical integration.

We have therefore revised the calculation using ERA5 surface pressure. At each grid cell and time step, pressure levels below the local surface pressure were excluded before calculating the vertically integrated moisture transport and moisture flux convergence. We retained 200 hPa as the upper boundary because it is well above the highest elevations of the Awash Basin (approximately 4300 m a.s.l., corresponding to roughly 600 hPa), and therefore encompasses the atmospheric column above the entire basin.

This correction substantially affects the resulting VIMC. The magnitude of the basin-mean convergence decreases by approximately one order of magnitude, from values of the order of 10^{12} to 10^{11} $\mathrm{m^3 year^{-1}}$, and the temporal variability is also modified. Importantly, the terrain-aware VIMC exhibits temporal variations that are more closely consistent with the observed variability of the lake surface areas. We have therefore replaced the original fixed-boundary calculation with the terrain-aware calculation in the revised manuscript and updated the corresponding results and discussion (Fig. 7).

Following the reviewer's suggestion, we also quantified the closure of the atmospheric moisture budget over the Awash Basin, using GLEAM as the evaporation estimate, CHIRPSv2 as the precipitation dataset, and ERA5 for both the vertically integrated moisture convergence and atmospheric moisture storage. We calculated the basin-integrated moisture budget as:

$$\frac{\partial W}{\partial t} = E_{\mathrm{basin}} - P_{\mathrm{basin}} + C \tag{1}$$

where (W) is the total column water vapor, P_{basin} is basin-averaged precipitation, E_{basin} is basin-averaged evapotranspiration, and (C) is the vertically integrated horizontal moisture flux convergence. All quantities were converted to monthly water-volume totals over the Awash Basin ($\mathrm{km}^3, \mathrm{month}^{-1}$).

The mean values over the common period are:

- **Evaporation (GLEAM):** $4.27 \times 10^9\ \mathrm{m}^3\,\mathrm{month}^{-1}$
- **Precipitation (CHIRPSv2):** $5.85 \times 10^9\ \mathrm{m}^3\,\mathrm{month}^{-1}$
- **Moisture convergence (ERA5):** $3.61 \times 10^9\ \mathrm{m}^3\,\mathrm{month}^{-1}$
- **Observed atmospheric storage change (dW/dt, ERA5):** $-1.13 \times 10^8\ \mathrm{m}^3\,\mathrm{month}^{-1}$
- $E - P + C$**:** $2.05 \times 10^9\ \mathrm{m}^3\,\mathrm{month}^{-1}$
- **Residual ($dW/dt - (E - P + C)$):** $2.05 \times 10^9\ \mathrm{m}^3\,\mathrm{month}^{-1}$

Thus, as anticipated, the atmospheric moisture budget does not close exactly. This is expected because precipitation, evaporation, vertically integrated moisture transport, and atmospheric moisture storage are derived from different datasets and because they have different spatial and temporal characteristics and associated uncertainties. Nevertheless, the three flux terms exhibit a strong and coherent seasonal cycle. In particular, precipitation, evaporation, and moisture convergence vary consistently through the year, with moisture convergence slightly preceding the precipitation maximum during approximately June–September.

To illustrate this behaviour, Fig. 8 shows the monthly time series of precipitation, evaporation, moisture convergence, (E-P+C), and (dW/dt) for 1981–1991, together with the corresponding monthly climatologies for the full 1981-2024 period. The figure also explicitly shows the residual between the atmospheric storage tendency and the estimated budget. This analysis demonstrates that, although exact closure is not achieved, the independently derived datasets capture a consistent seasonal evolution of the atmospheric moisture budget over the Awash Basin.

Reviewer comment: *"Sect. 2.3.3: Given your precipitation-based ITCZ identification, please consider replacing the term ITCZ with a more appropriate term. Please see discussions in: Nicholson (2009; 2018)."*

Response

We thank the reviewer for pointing this out. We agree that the term ITCZ is not the most appropriate terminology for our precipitation-based diagnostic. As discussed by Nicholson (2009); Nicholson et al. (2018), over Africa the surface convergence zone and the rainfall maximum are not necessarily co-located, and the term ITCZ is therefore better avoided when referring specifically to the precipitation maximum.

We have accordingly replaced ITCZ with tropical rain belt throughout Sect. 2.3.3 and now refer to the precipitation-weighted mean latitude as a measure of the latitudinal position of the regional tropical rain belt. This terminology more accurately reflects

what is diagnosed by our precipitation-based method while remaining consistent with the terminology used by Nicholson (2009); Nicholson et al. (2018).

Reviewer comment: *"L263–264: How can you tell this is the driver? If that's the case, why isn't it the same across the other parts of the time series?"*

Response

We thank the reviewer for raising this point. We agree that our original wording was too strong in attributing the underestimation of lake expansion in 2024 to the monthly time step of GR2M and its performance during MAM. This explanation was initially proposed as a hypothesis to account for the increasing discrepancy between simulated runoff and observed lake-surface dynamics after 2020. In the revised manuscript, we investigated this hypothesis further by examining the temporal characteristics of extreme hydrological events in the Awash Basin. We found no significant trends in their frequency or duration, and therefore no observational evidence supporting changes in these characteristics as an explanation for the increasing discrepancy.

We subsequently explored the potential influence of anthropogenic water management, particularly the operation of Tendaho Reservoir, as an alternative explanation. As detailed in our response to a previous general comment (line 225), the Landsat-based analysis reveals changes in the temporal relationships between Tendaho Reservoir, Afambo–Gemeri, and Lake Abhe, consistent with an increasing influence of reservoir regulation on downstream lake dynamics. Although these observations do not allow us to directly identify or quantify individual reservoir releases, they provide a plausible explanation for why the discrepancy between simulated runoff and observed lake-surface dynamics varies over time. GR2M does not explicitly represent reservoir operations, controlled releases, or changes in water withdrawals, and therefore cannot capture their potential effects on downstream lake levels.

We have consequently removed the original attribution to the GR2M monthly time step and MAM performance as specific causes of the 2024 discrepancy. The revised discussion distinguishes the initial hypothesis, which is not supported by the additional evidence, from the potential influence of changing anthropogenic regulation on the downstream lake system. The corresponding discussion has been revised in the new Appendix B.

Reviewer comment: *"L280–282: I'm not sure I can see this. Excluding the period of 1985–1990, they seem to be roughly static. Can you explain how you got to this conclusion? The same holds for the conclusion in 285–287. Please quantify/explain."*

Response

We thank the reviewer for pointing this out. We agree that our original wording was too strong and that the changes are not sufficiently apparent from the time series alone. We have therefore quantified the differences between the periods before and after 2000 and revised the text accordingly.

For actual evapotranspiration, we calculated the mean annual totals for the two periods for each dataset and added the resulting distributions as a new supplementary figure (Figures 9 and 10). Four of the five datasets show lower mean annual evapotranspiration after 2000:

- GLEAM decreases from 54.91×10^9 to 48.38×10^9 $\mathrm{m^3\,year^{-1}}$ (-11.9%);
- PCR-GLOBWB from 52.66×10^9 to 45.70×10^9 $\mathrm{m^3\,year^{-1}}$ (-13.2%);
- HTESSEL from 65.43×10^9 to 58.63×10^9 $\mathrm{m^3\,year^{-1}}$ (-10.4%);
- LISFLOOD from 56.11×10^9 to 52.98×10^9 $\mathrm{m^3\,year^{-1}}$ (-5.6%).

In contrast, W-GAP shows only a small difference, decreasing from 80.41×10^9 to 79.01×10^9 $\mathrm{m^3\,year^{-1}}$ (-1.7%).

Thus, rather than characterizing the change as uniformly "considerable," we now describe it as a reduction in annual evapotranspiration that is consistent across four of the five datasets, with a magnitude of approximately 6–13 %, while noting the much smaller change in W-GAP.

To quantify the difference between the two periods, we compared the mean synchronicity index before and after 2000. All four combinations show a reduction after 2000:

- CHIRPS–GLEAM decreases from 350.20 $\mathrm{mm\,year^{-1}}$ during 1981–1999 to 338.24 $\mathrm{mm\,year^{-1}}$ during 2000–2024 (-3.4%);
- CHIRPS–W-GAP decreases from 300.92 $\mathrm{mm\,year^{-1}}$ to 293.31 $\mathrm{mm\,year^{-1}}$ during 2000–2016 (-2.5%);
- CHIRPS–PCR-GLOBWB decreases from 284.38 $\mathrm{mm\,year^{-1}}$ to 274.35 $\mathrm{mm\,year^{-1}}$ during 2000–2014 (-3.5%);
- CHIRPS–CRU decreases from 339.55 $\mathrm{mm\,year^{-1}}$ to 332.00 $\mathrm{mm\,year^{-1}}$ during 2000–2024 (-2.2%).

Thus, although the magnitude of the difference is modest, its consistency across all four datasets indicates a shift toward weaker rainfall–evapotranspiration synchronisation after 2000.

We have consequently revised the text to avoid implying a progressive or monotonic decline and now describe the result as a shift between periods. We have also added a supplementary figure showing the distributions and individual annual values for the evapotranspiration and synchronicity index datasets, which makes the differences between the pre- and post-2000 periods more transparent.

Reviewer comment: *"L289–291: Are these trends similar to gauge-based trends? How about other dataset? This question rises whenever a long-term satellite product is used, because sensors and satellites are being replaced, and it is hard to assess the validity of these trends."*

Response

We thank the reviewer for raising this important point regarding the robustness of long-term precipitation trends derived from satellite-based products and the need to compare them with in-situ observations and alternative gridded datasets.

In-situ precipitation observations are scarce and difficult to access in the Awash Basin. We identified available station records in SIEREM (Système d'Information Environnementale sur les Ressources en Eau et leur Modélisation), a hydrometeorological database assembled by HydroSciences Montpellier (http://www.hydrosciences.fr/sierem/). Within the Awash Basin, we identified five stations with more than 15 years of observations after 1980. Although these records are too short to provide a robust quantitative validation of trends over the full 1981–2024 period considered in our study, they allow a qualitative assessment of the consistency of seasonal trends. We therefore calculated trends in March–May (MAM) and July–September (JAS) precipitation at these stations (Fig. 11).

During JAS, four of the five stations exhibit positive trends, ranging from 4.47 to 19.33 $\mathrm{mm\,year^{-1}}$, although none is statistically significant. During MAM, three stations exhibit negative trends, one exhibits a positive trend, and one shows no trend; none of these trends is statistically significant. These results are qualitatively consistent with the spatial patterns derived from CHIRPS, which indicate generally increasing JAS precipitation across much of the basin, except for parts of the central basin, and decreasing MAM precipitation across most of the basin. However, given the short station records and the associated statistical uncertainty, these comparisons should be interpreted as an indication of broad consistency rather than as a quantitative validation of the gridded trends.

We also clarify that CHIRPS v2.0 is not a purely satellite-based product: it combines satellite-derived precipitation estimates with in-situ rain-gauge observations. Nevertheless, we agree with the reviewer that long-term gridded precipitation trends should be evaluated against alternative products, particularly given potential uncertainties associated with changes in satellite sensors, observing systems, and data integration procedures.

To address this concern, we additionally examined precipitation trends in four gridded datasets: CRU TS v4.09, CenTrends, GPCP, and GPCC (Fig. 12). CRU TS and CenTrends are based on in-situ station observations, while GPCC is also gauge-based and GPCP combines satellite and gauge observations. The comparison reveals broadly consistent large-scale seasonal patterns across the products: all four indicate increasing JAS precipitation over the lower Awash Basin and generally decreasing MAM precipitation across most of the basin. An exception is CRU TS, which indicates positive MAM trends over the upper basin and negative trends over the lower basin. These agreements provide additional evidence that the principal seasonal patterns identified in our analysis are not unique to CHIRPS. However, because the products may share some underlying gauge observations and differ in their spatial and temporal coverage and interpolation methods, this agreement should not be interpreted as fully independent validation.

We also considered CMORPH and TRMM 3B42 but did not include them in this comparison because their shorter temporal coverage is less suitable for assessing trends over the period examined in our study. Among the alternative datasets, we selected CenTrends for inclusion in the revised analysis because its gauge-based observations, long temporal coverage, and relatively high spatial resolution make it particularly suitable for examining long-term precipitation trends in the Awash Basin. The other datasets are used as supplementary robustness checks rather than incorporated into the main analysis, given their respective limitations in temporal coverage and/or spatial resolution for our specific objectives.

We have therefore incorporated the CenTrends comparison into the revised analysis and clarified the observational basis and limitations of the precipitation products considered. We believe these additions strengthen the assessment of the robustness of the identified seasonal trends while acknowledging the limitations imposed by the scarcity of long-term in-situ observations in the region.

Reviewer comment: *"L299: Can you explain what is this warming? Isn't it also the human-induced warming?"*

Response

Thank you for pointing out this ambiguity. We agree that the origin of the warming should be clarified. Hagos et al. (2024) investigate the influence of northwestern Eurasian land-surface warming on East African MAM rainfall through its effects on the regional Hadley circulation and moisture transport. However, they do not explicitly attribute this regional warming to anthropogenic forcing. Instead, they discuss the potential role of SST variability, particularly patterns associated with the PDO and La Niña-like conditions, as possible drivers of the regional warming.

We have therefore revised the sentence to distinguish the atmospheric mechanism identified by Hagos et al. from the attribution of the regional warming itself. We also retain the separate interpretation of Tierney et al. (2015), who suggest that the unusual twentieth-century drying in the Horn of Africa may have an anthropogenic component based on paleoclimate evidence.

Reviewer comment: *"L316–327: I suggest avoiding the interpretation of monthly mean 500-hPa vertical velocity as a direct proxy for convective activity. At ERA5 resolution, and particularly after monthly averaging, w500 primarily represents large-scale mean ascent/subsidence rather than convective updraughts. Indeed, the stated motivation for monthly averaging: to suppress short-lived convective variability, appears inconsistent with using the resulting field to characterise convection itself. I suggest describing this variable as an indicator of large-scale vertical motion or the large-scale dynamical environment associated with convection, unless its relationship with an independent measure of convective activity is demonstrated."*

Response

Thank you for this important clarification. We agree that monthly mean ERA5 vertical velocity at 500 hPa (ω_{500}) should not be interpreted as a direct proxy for convective activity. At this temporal resolution, the variable primarily characterizes persistent large-scale atmospheric ascent and subsidence, rather than individual convective updrafts.

We have therefore revised the terminology throughout the manuscript to describe ω_{500} as an indicator of large-scale vertical motion and the associated dynamical environment favouring or inhibiting convection. We have also revised the explanation of monthly averaging to clarify that its purpose is to emphasize persistent large-scale circulation patterns while reducing the influence of short-lived atmospheric fluctuations. We have accordingly adjusted the interpretation of the results to avoid equating large-scale ascent directly with convective activity.

Reviewer comment: *"Fig. 6: Please add the basin shape to this figure. With regards to Fig. 6b, I think detrending is necessary, as we expect trends (in both datasets) to have an impact on this result."*

Response

Thank you for this suggestion. We agree that removing long-term trends is important to ensure that the correlations reflect covariability beyond the underlying long-term changes in the two datasets. In fact, linear detrending was already applied prior to the correlation analysis, both to the Lake Abhe surface-area time series and independently to the SST time series at each grid cell. However, we acknowledge that this procedure was not sufficiently explicit in the original figure caption. We have therefore revised the caption to clarify that both datasets were detrended before calculating the correlations. We have also added the outline of the Awash Basin to Figure 6a, as suggested.

Reviewer comment: *"Fig. 7: Why Lake Abhe only? Why not showing the observed lake time series too?"*

Response

Thank you for this suggestion. We initially focused on Lake Abhe because it is the largest of the three lakes and contributes substantially to the variability of the combined lake-system surface area. Regarding the observational time series, we initially did not include it because continuous observations are only available from 1999 onwards. This relatively short record limits the robustness of correlation analyses between observed lake-system surface area and SST anomalies, particularly when investigating potential relationships with large-scale climate variability. Nevertheless, we agree that showing the observations alongside the simulations would provide a useful visual comparison over the period of data availability. We have therefore revised Figure 7 to show the combined surface area of Lakes Afambo, Gemeri and Abhe, together with the available observational estimates.

Reviewer comment: *"L356: How do you define the "boreal summer" is it the same as your long-rains definition?"*

Response

Thank you for pointing out this ambiguity. In this study, "boreal summer" refers specifically to July–August–September (JAS). We agree that the original terminology was insufficiently precise and have replaced "boreal summer" with "July–August–September (JAS)" in the manuscript.

Reviewer comment: *"L412: What is the meaning of "enhancing kinetic energy"? Shouldn't the interaction with topography reduce the speed and kinetic energy of the flow?"*

Response

Thank you for raising this point. We agree that our original wording was potentially misleading, as orographic interaction can also lead to flow deceleration through frictional and drag processes. However, Masiwal et al. (2023) identify localized kinetic-energy generation over the East African highlands, associated with the interaction between the low-level flow and orography.

This does not imply a systematic acceleration of the flow, as kinetic-energy generation, advection and dissipation act simultaneously. We have therefore revised the sentence to clarify that East African orography modifies the low-level wind structure and contributes to localized kinetic-energy generation over the highlands, rather than suggesting a general enhancement of kinetic energy throughout the flow.

Reviewer comment: *"Fig. 12e: Can you please add the 0-precipitation trend line? If all values are positive, you could put it in the lowermost part of the axis, but otherwise it's hard to distinguish negative from positive trends."*

Response

Thank you for this suggestion. We agree that including a zero-trend reference line would make it easier to distinguish positive and negative precipitation trends. We have therefore added a horizontal line indicating zero trend to Fig. 12e, improving the readability of the figure.

Reviewer comment: *"L434: How can you tell that this increase in specific humidity comes from increased orographic precipitation? Can it also be there because of advection?"*

Response

Thank you for raising this important point. We agree that the increase in mid-tropospheric specific humidity cannot be attributed exclusively to orographic precipitation based on the diagnostics presented. Enhanced horizontal moisture advection associated with the strengthening Somali Jet may also contribute to the observed humidity increase. We have therefore revised the text to distinguish the enhanced moisture transport from the potential role of orographic lifting in promoting precipitation over the Somali Plateau. We now acknowledge that the relative contributions of horizontal moisture advection and orographic processes cannot be quantified from the analyses presented.

Reviewer comment: *"All figures presenting maps from -25 to 25: Consider reducing the latitudinal span of these figures. Perhaps 40W to 90E or something like this."*

Response

Thank you for this suggestion. We agree that the geographical extent of these maps can be reduced to improve their readability and focus on the regions most relevant to our analysis. We have therefore revised the geographical domains of the corresponding figures to provide a more focused representation of the relevant regions.

Reviewer comment: *"Fig. 13: Can you add the Sahara and Sahel rectangles related to Fig. 13b to one of the maps in this figure? Can you add political borders to panel c and d or zoom out a bit so its easier to read these maps? And the river catchment as well?"*

Response

Thank you for these helpful suggestions. We have revised Fig. 13 to improve its geographical context and readability. The Sahara and Sahel rectangles used in Fig. 13b have been added to the corresponding spatial map to indicate the regions considered in the analysis. We have also included political boundaries and the Awash Basin outline in panels c and d, and adjusted the map extent where necessary to facilitate the identification of the geographical features and improve readability.

Reviewer comment: *"L484: I'm not sure I understand what the tool is. Is the problem stated common throughout the globe for pre-1998 data?"*

Response

Thank you for pointing out this lack of clarity. We agree that the expression "transferable tool" was too vague and that our original wording could be interpreted as suggesting that the pre-1998 observational gap is a global issue. Our statement specifically refers to the limited availability of long-term lake observations for the Awash terminal lake system. We have revised the text to clarify that the proposed approach consists of coupling a rainfall–runoff model with a lake water-balance model, and that this framework may be applicable to other data-scarce lake basins, depending on data availability and local conditions.

Reviewer comment: *"Technical corrections"*

Response

We have also reviewed and implemented all the technical corrections and minor editorial changes suggested by the reviewer throughout the manuscript.

References

Armon, M., Dente, E., Shmilovitz, Y., Mushkin, A., Cohen, T. J., Morin, E., and Enzel, Y.: Determining bathymetry of shallow and ephemeral desert lakes using satellite imagery and altimetry, Geophysical Research Letters, 47, e2020GL087 367, 2020.

Awaleh, M. O., Hoch, F. B., Boschetti, T., Soubaneh, Y. D., Egueh, N. M., Elmi, S. A., Mohamed, J., and Khaireh, M. A.: The geothermal resources of the Republic of Djibouti—II: Geochemical study of the Lake Abhe geothermal field, Journal of Geochemical Exploration, 159, 129–147, 2015.

Billi, P., Alemu, Y. T., and Ciampalini, R.: Increased frequency of flash floods in Dire Dawa, Ethiopia: Change in rainfall intensity or human impact?, Nat. Hazards, 76, 1373–1394, https://doi.org/10.1007/s11069-014-1554-0, 2015.

Chen, X., Workneh, H. T., Ma, Y., Tegegne, E. B., and Kassahun, A.: Quantifying Runoff and Total River Discharge in Ethiopia Based on Water Balance Using Satellite and Reanalysis Datasets, SSRN, https://doi.org/10.2139/ssrn.5010515, preprint, SSRN, 2024.

Coron, L., Thirel, G., Delaigue, O., Perrin, C., and Andréassian, V.: The suite of lumped GR hydrological models in an R package, Environmental Modelling & Software, 94, 166–171, https://doi.org/https://doi.org/10.1016/j.envsoft.2017.05.002, 2017.

Crétaux, J.-F. and Birkett, C.: Lake studies from satellite radar altimetry, Comptes Rendus Geoscience, 338, 1098–1112, https://doi.org/https://doi.org/10.1016/j.crte.2006.08.002, la Terre observée depuis l'espace, 2006.

Crétaux, J.-F., Arsen, A., Calmant, S., Kouraev, A., Vuglinski, V., Bergé-Nguyen, M., Gennero, M.-C., Nino, F., Abarca Del Rio, R., Cazenave, A., and Maisongrande, P.: SOLS: A lake database to monitor in the Near Real Time water level and storage variations from remote sensing data, Advances in Space Research, 47, 1497–1507, https://doi.org/https://doi.org/10.1016/j.asr.2011.01.004, 2011.

Dinku, T., Funk, C., Peterson, P., Maidment, R., Tadesse, T., Gadain, H., and Ceccato, P.: Validation of the CHIRPS satellite rainfall estimates over eastern Africa, Q. J. R. Meteorol. Soc., 144, 292–312, https://doi.org/10.1002/qj.3244, 2018.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, Journal of Hydrology, 377, 80–91, https://doi.org/https://doi.org/10.1016/j.jhydrol.2009.08.003, 2009.

Hagos, S., Chang, C.-C. J., Ma, P.-L., Lubis, S. W., Balaguru, K., Shi, P., et al.: Synchronization of the recent decline of East African long rains and northwestern Eurasian warming, J. Geophys. Res. Atmos., 129, e2024JD041 033, https://doi.org/10.1029/2024JD041033, 2024.

Huang, C. H., Zhang, S., Shah, D., Yadav, A., Li, Y., Zhao, G., and Gao, H.: 3D-LAKES: Three-Dimensional Global Lake and Reservoir Bathymetry from ICESat-2 Altimetry and Landsat Imagery, Scientific Data, 12, 1625, https://doi.org/10.1038/s41597-025-05911-y, 2025.

Kebede, S., Charles, K., Godfrey, S., MacDonald, A., and Taylor, R. G.: Regional-scale interactions between groundwater and surface water under changing aridity: evidence from the River Awash Basin, Ethiopia, Hydrol. Sci. J., 66, 450–463, https://doi.org/10.1080/02626667.2021.1874613, 2021.

Knighton, A. D. and Nanson, G. C.: Flow transmission along an arid zone anastomosing river, Cooper Creek, Australia, Hydrological Processes, 8, 137–154, 1994.

Morin, E., Grodek, T., Dahan, O., Benito, G., Kulls, C., Jacoby, Y., Van Langenhove, G., Seely, M., and Enzel, Y.: Flood routing and alluvial aquifer recharge along the ephemeral arid Kuiseb River, Namibia, Journal of Hydrology, 368, 262–275, 2009.

Nicholson, S. E.: A revised picture of the structure of the "monsoon" and land ITCZ over West Africa, Climate Dynamics, 32, 1155–1171, 2009.

Nicholson, S. E.: Climate and climatic variability of rainfall over eastern Africa, Rev. Geophys., 55, 590–635, https://doi.org/10.1002/2016RG000544, 2017.

Nicholson, S. E., Funk, C., and Fink, A. H.: Rainfall over the African continent from the 19th through the 21st century, Global and planetary change, 165, 114–127, 2018.

Palmer, P. I., Wainwright, C. M., Dong, B., Maidment, R. I., Wheeler, K. G., Gedney, N., Hickman, J. E., Madani, N., Folwell, S. S., Abdo, G., et al.: Drivers and impacts of Eastern African rainfall variability, Nat. Rev. Earth Environ., 4, 254–270, https://doi.org/10.1038/s43017-023-00397-x, 2023.

Ricko, M., Carton, J. A., Birkett, C. M., and Cretaux, J.-F.: Intercomparison and validation of continental water level products derived from satellite radar altimetry, Journal of Applied Remote Sensing, 6, 061 710, https://doi.org/10.1117/1.JRS.6.061710, 2012.

Tierney, J. E., Ummenhofer, C. C., and DeMenocal, P. B.: Past and future rainfall in the Horn of Africa, Sci. Adv., 1, e1500 682, https://doi.org/10.1126/sciadv.1500682, 2015.

Wolde-Georgis, T.: El Niño and Drought Early Warning in Ethiopia, https://ssrn.com/abstract=1589710, 1997.

Woldeyohanis, B. A.: Operation Modeling For Tendaho Reservoir Using HEC-ResSim, Thèse de doctorat, Addis Ababa University, 2010.

Zhao, G., Li, Y., Zhou, L., et al.: Evaporative water loss of 1.42 million global lakes, Nat. Commun., 13, 3686, https://doi.org/10.1038/s41467-022-31125-6, 2022.

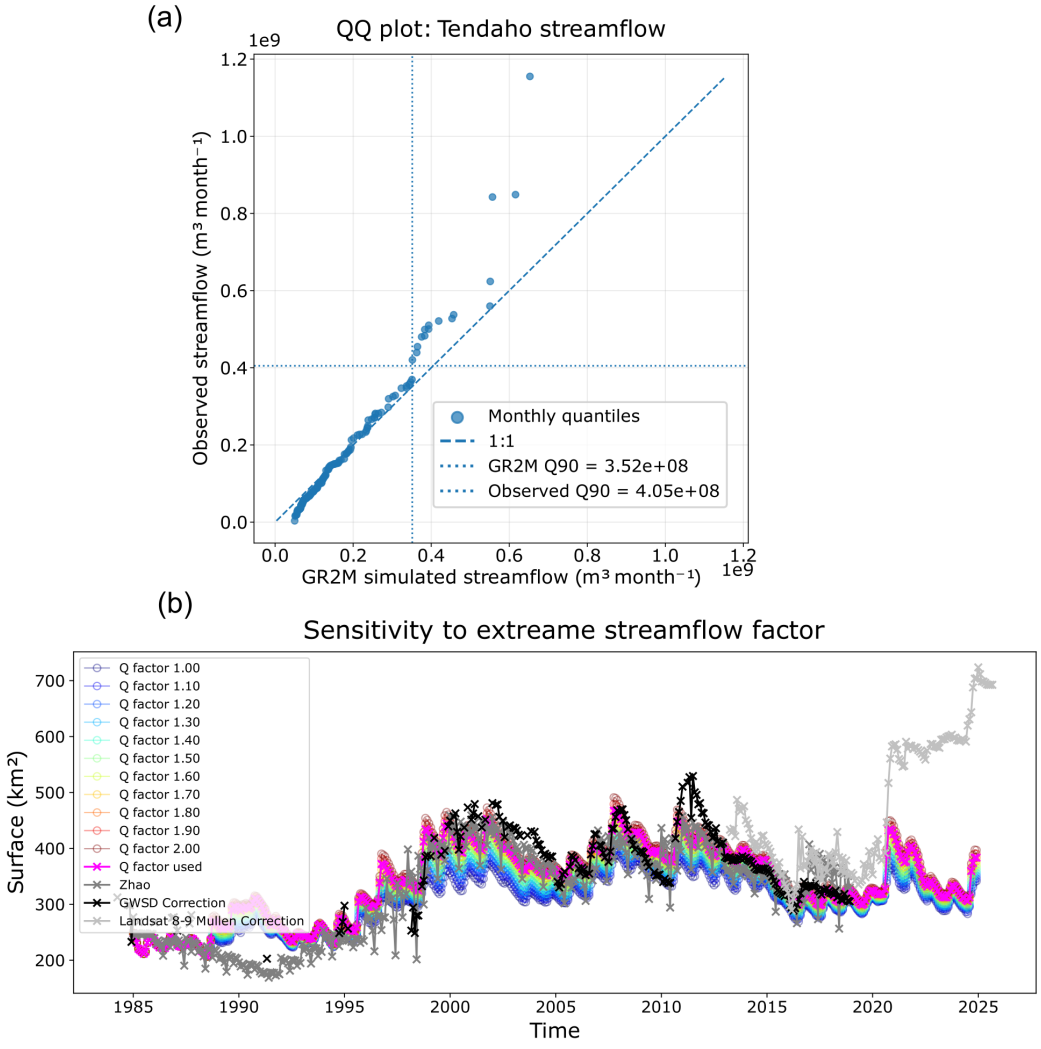


Figure 1. (a) Quantile–quantile plot comparing observed discharge at Tendaho from the Global Runoff Data Centre (GRDC) database with discharge simulated using the GR2M model forced with CHIRPS v2 precipitation and GLEAM v4.2a potential evapotranspiration data. The 90th percentiles of the observed and simulated discharge distributions are indicated by vertical and horizontal dashed lines, respectively. (b) Time series of the total surface area of the Afambo, Gemeri, and Abhe lakes, showing satellite-derived estimates from the Global Surface Water Dynamics (GSWD) dataset (black line) and Landsat 8–9 imagery (light gray line), together with the reconstruction from Zhao et al. (2022) (dark gray line). Lake-model simulations obtained using GR2M-simulated streamflow are also shown for scaling factors ranging from 1 to 2, applied to flows exceeding the 90th percentile. The magenta line indicates the simulation using the selected high-flow scaling factor of 1.8.

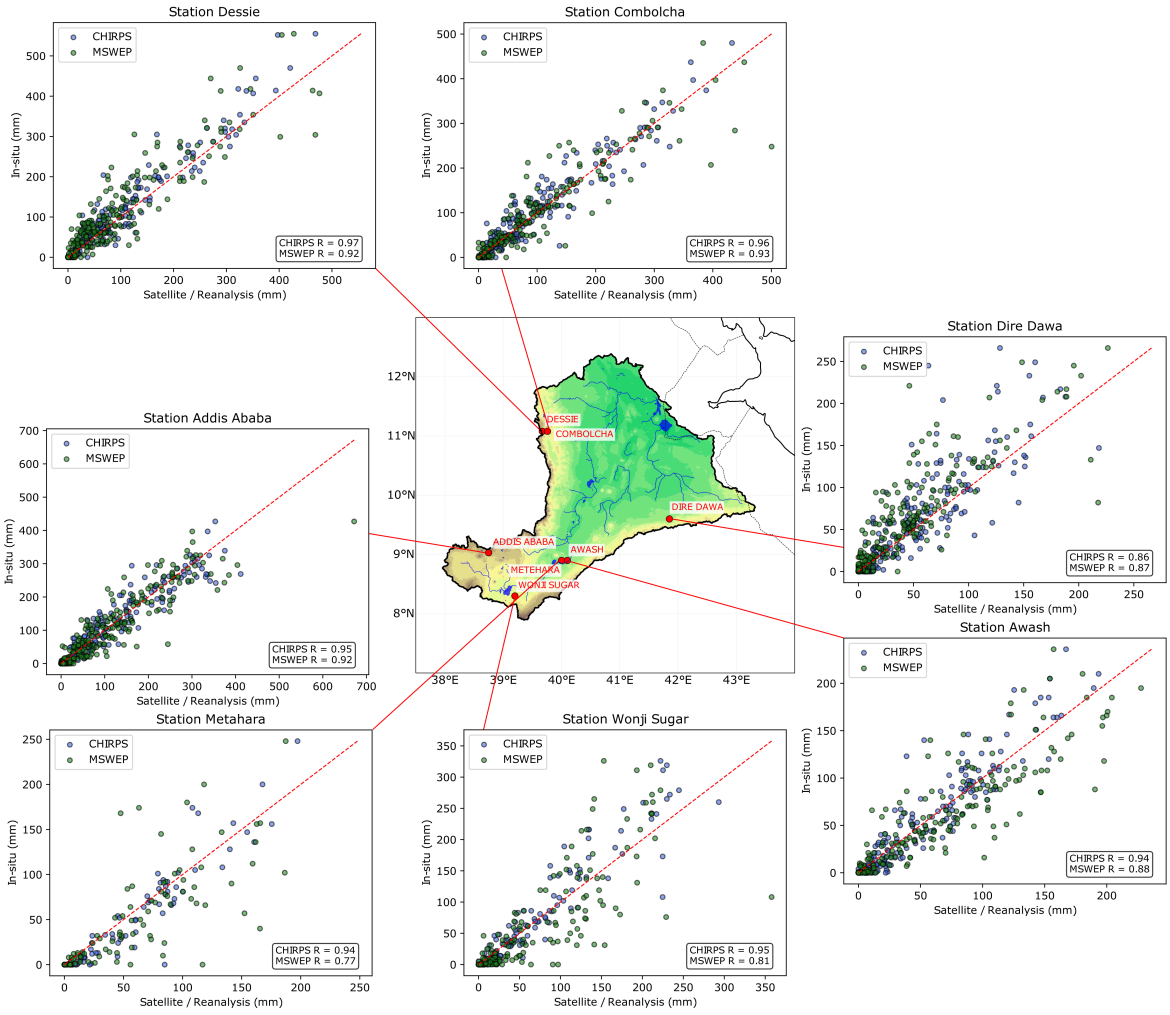


Figure 2. Map showing the locations of weather stations available in the SIEREM database (Système d'Informations Environnementales sur les Ressources en Eau et leur Modélisation) and comparison of monthly accumulated precipitation from station observations, CHIRPSv2, and MSWEP.

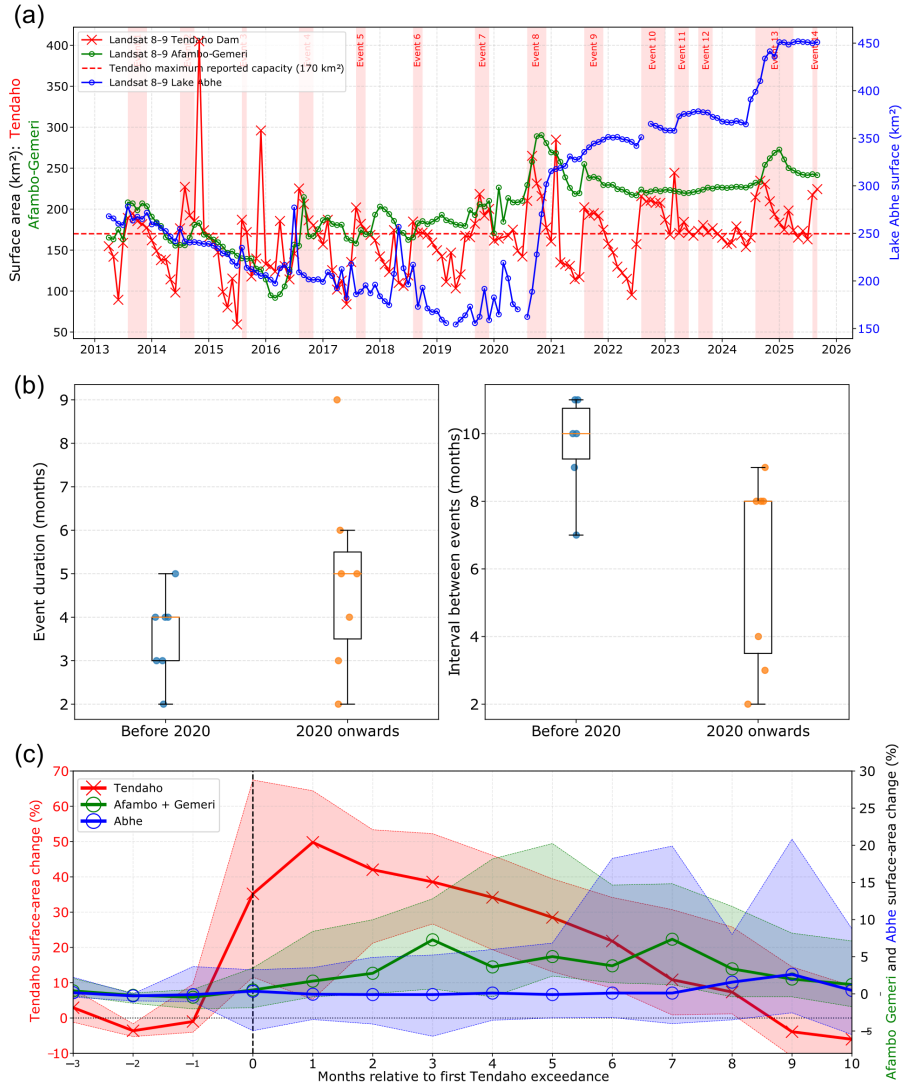


Figure 3. (a) Time series of water surface area derived from Landsat 8–9 satellite imagery for the Afambo–Gemeri lake complex, Lake Abhe, and the Tendaho Reservoir, located upstream of the Afambo–Gemeri–Abhe lake system. The horizontal dashed line indicates the maximum design capacity of the Tendaho Reservoir (170 km^2), as reported by Woldeyohanis (2010). Isolated, anomalous spikes in the Tendaho time series, observed toward the end of 2014 and 2015 and at the beginning of 2021, were excluded from the analysis because they likely reflect contamination or other artefacts in the satellite imagery rather than actual changes in reservoir surface area. (b) Box plots showing the duration of capacity-exceedance events (defined as consecutive months during which the Tendaho Reservoir surface area exceeded 170 km^2) and the interval between consecutive events, separated into periods before and after 2020. (c) Composite analysis of events during which the surface area of the Tendaho Reservoir exceeded its maximum design capacity. Surface areas are shown from 3 months before to 10 months after the onset of each event. Solid lines represent the median across events, and shaded areas indicate the interquartile range (25th–75th percentiles). For each event, surface areas were normalized by dividing them by the mean surface area of the corresponding water body during the 3 months preceding the event, allowing relative changes to be compared across events.

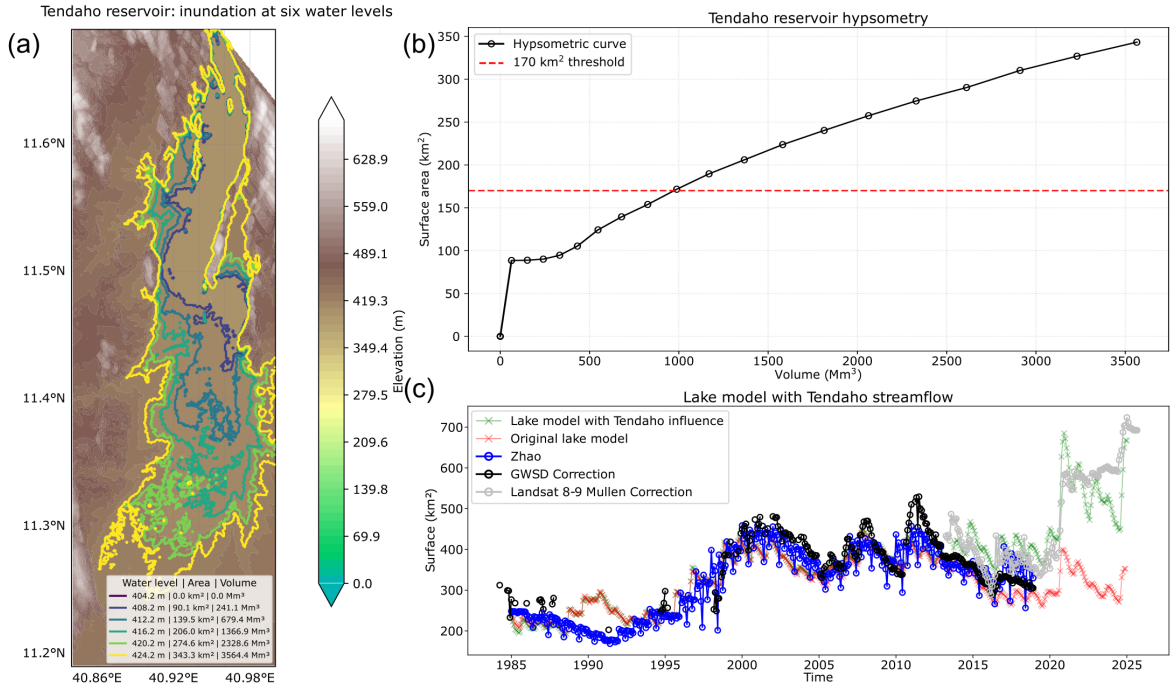


Figure 4. (a) SRTM-derived topography of the Tendaho Dam area (shaded; elevation in m above sea level), used as a proxy for the pre-dam terrain and reservoir bathymetry. Contour lines indicate successive water levels, defined by elevation increments relative to the minimum elevation at the dam outlet. (b) Hypsometric relationship derived from the SRTM topography, relating reservoir water level to inundated surface area (km^2) and estimated storage volume (Mm^3). The red horizontal line indicates the maximum design water level reported by Woldeyohanis (2010). (c) Combined surface area of the Afambo–Gemeri–Abhe lake system (km^2), derived from the corrected Global Surface Water Dataset (GSWD), extended to 2025 using Landsat 8–9 imagery, the reconstruction of Zhao et al. (2022), and GR2M–lake model simulations without and with additional water inputs representing the potential influence of Tendaho Reservoir.

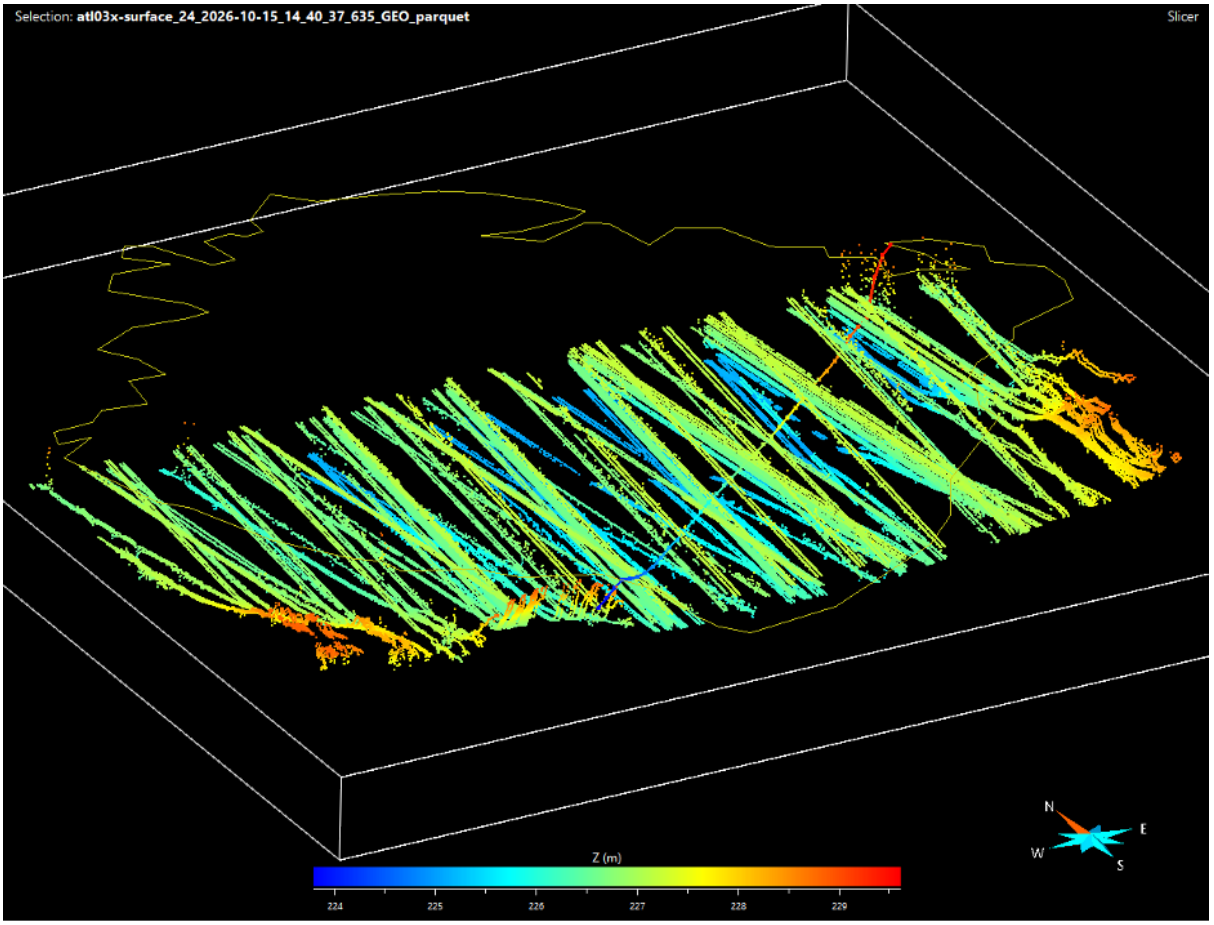


Figure 5. Outline of Lake Abhe and locations of the ICESat-2 ATL03 profiles used to validate the bathymetry employed in this study.

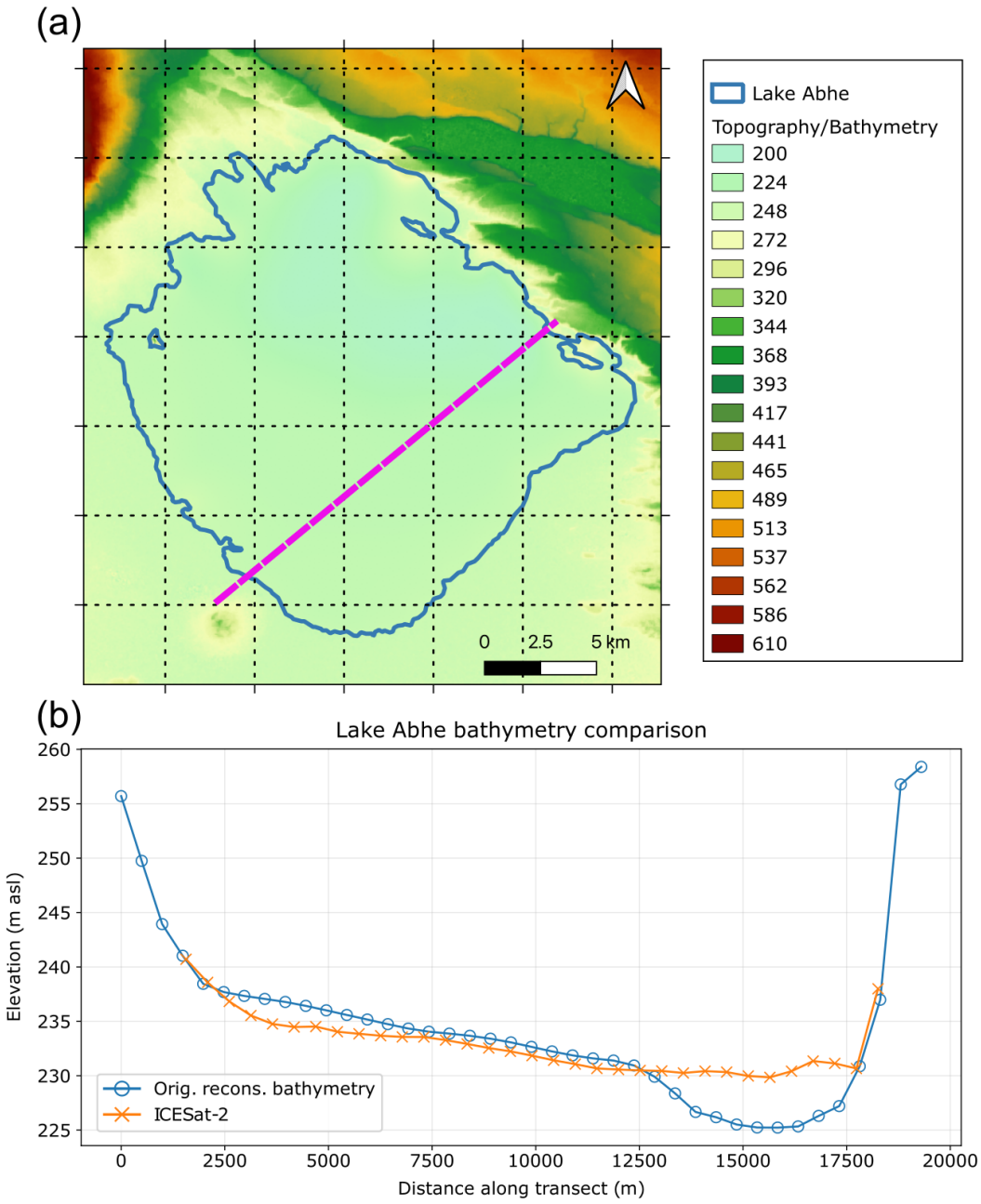


Figure 6. (a) Topography and bathymetry of Lake Abhe and its surroundings used in this study. Elevations are expressed in metres above sea level. The lake outline is derived from the HydroLAKES dataset, and the magenta line indicates the location of the cross-section shown in panel (b). (b) Comparison of lake-floor elevations along the selected cross-section, based on the bathymetric reconstruction developed in this study and elevation estimates derived from ICESat-2 ATL03 observations.

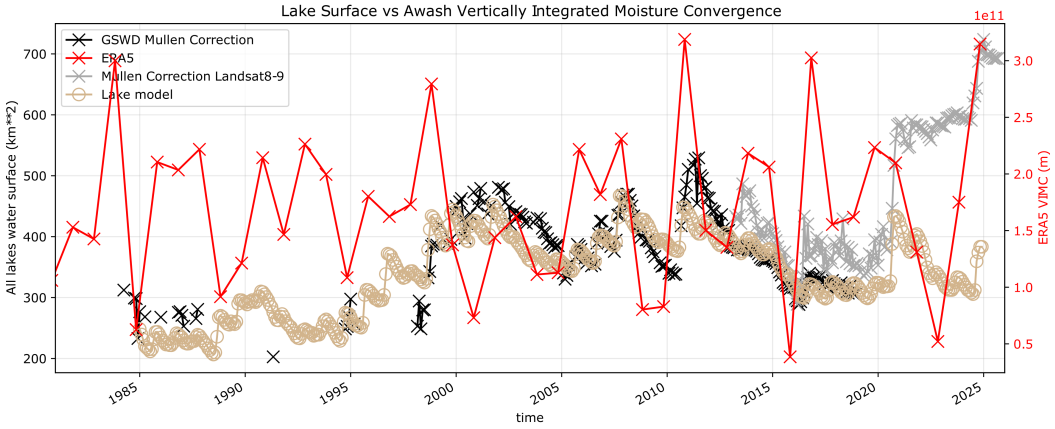


Figure 7. Time series of the combined surface area of the Afambo, Gemeri, and Abhe lakes, based on GSWD, Landsat 8–9 observations, and lake model simulations, and vertically integrated atmospheric moisture convergence in the Awash Basin derived from ERA5 data. Values are aggregated by hydrological year (October–September). This figure replaces the corresponding panel (Fig. 4d) in the original manuscript.

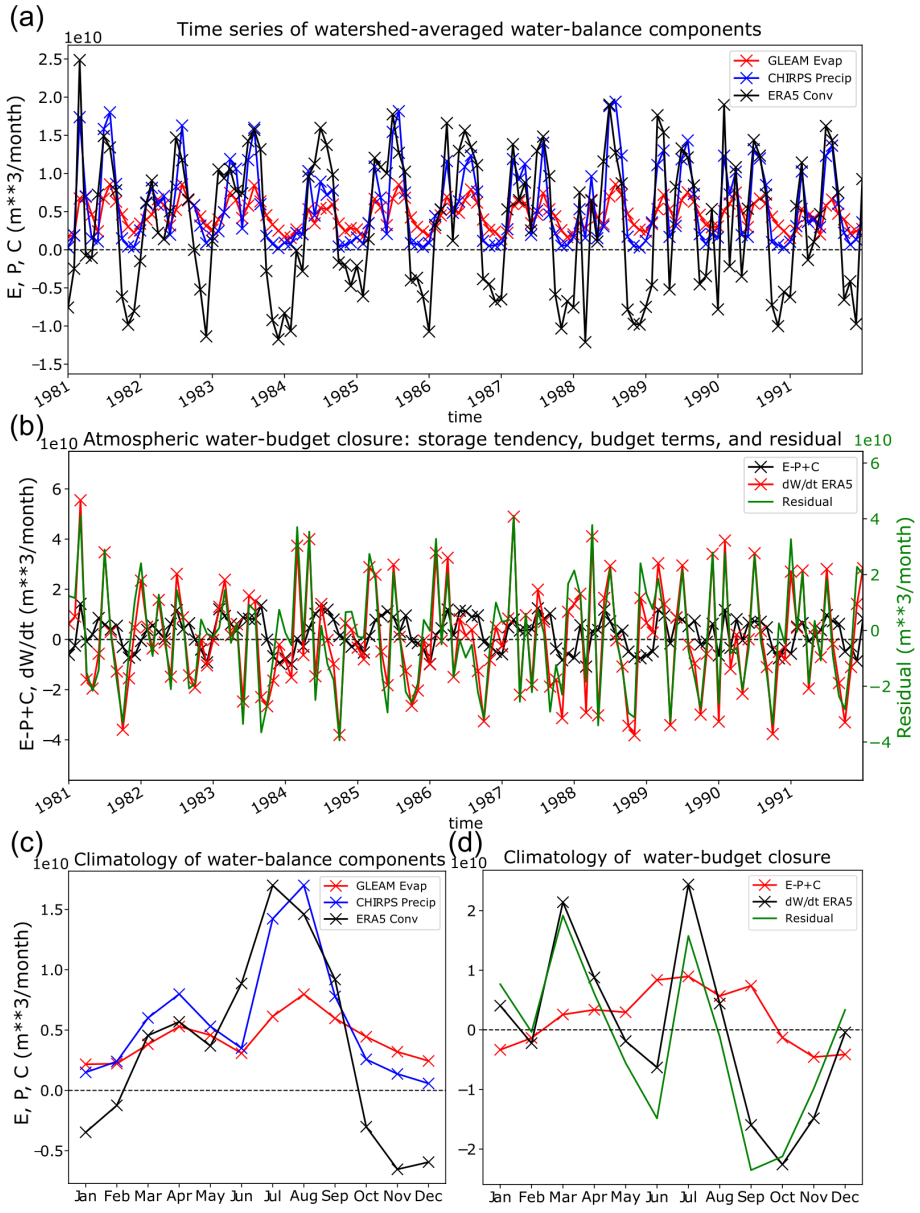


Figure 8. (a) Monthly time series of the atmospheric water-balance components over the Awash Basin: precipitation (P) from CHIRPSv2, actual evapotranspiration (E) from GLEAMv4.2a, and vertically integrated atmospheric moisture-flux convergence (C) from ERA5. The 1981–1991 period is shown to facilitate comparison of the temporal evolution of these components. (b) Monthly time series of the tendency of vertically integrated atmospheric moisture storage (dW/dt), the combined terms $E - P + C$, and the residual, defined as $dW/dt - (E - P + C)$. (c) Climatological monthly means of the atmospheric water-balance components over 1981–2024. (d) Climatological monthly means of dW/dt, $E - P + C$, and the residual over the same period.

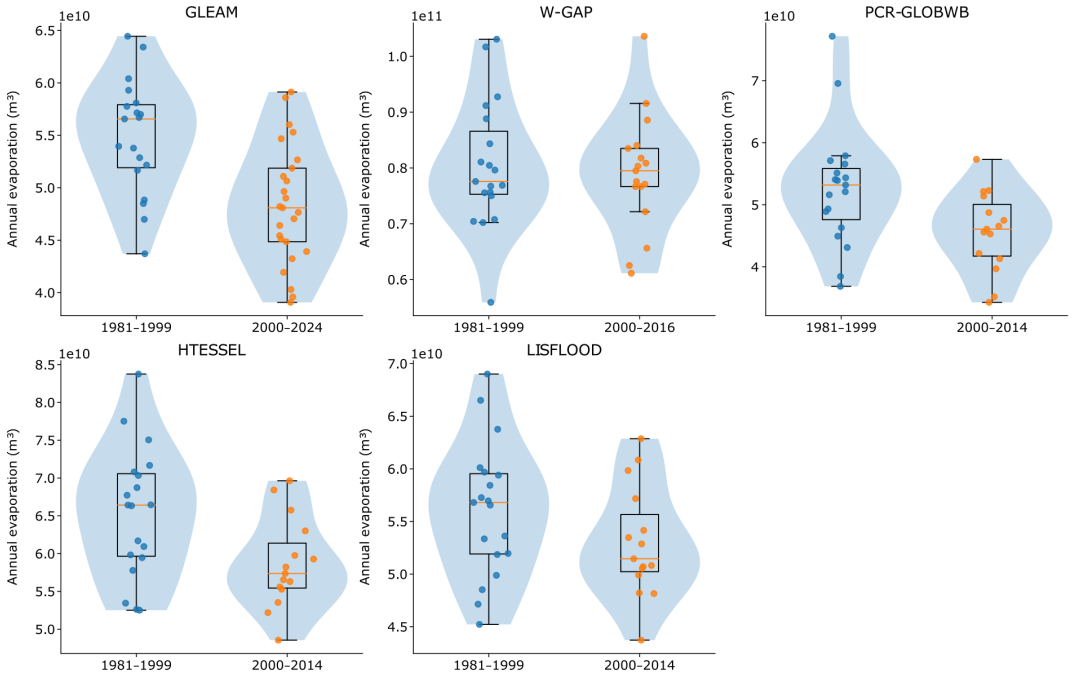


Figure 9. Violin and box plots of total annual evaporation m^3 for the October–September hydrological year, derived from five evaporation datasets (GLEAM v4.2a, WaterGAP3, PCR-GLOBWB, HTESSEL, and LISFLOOD), for the periods before and after 2000.

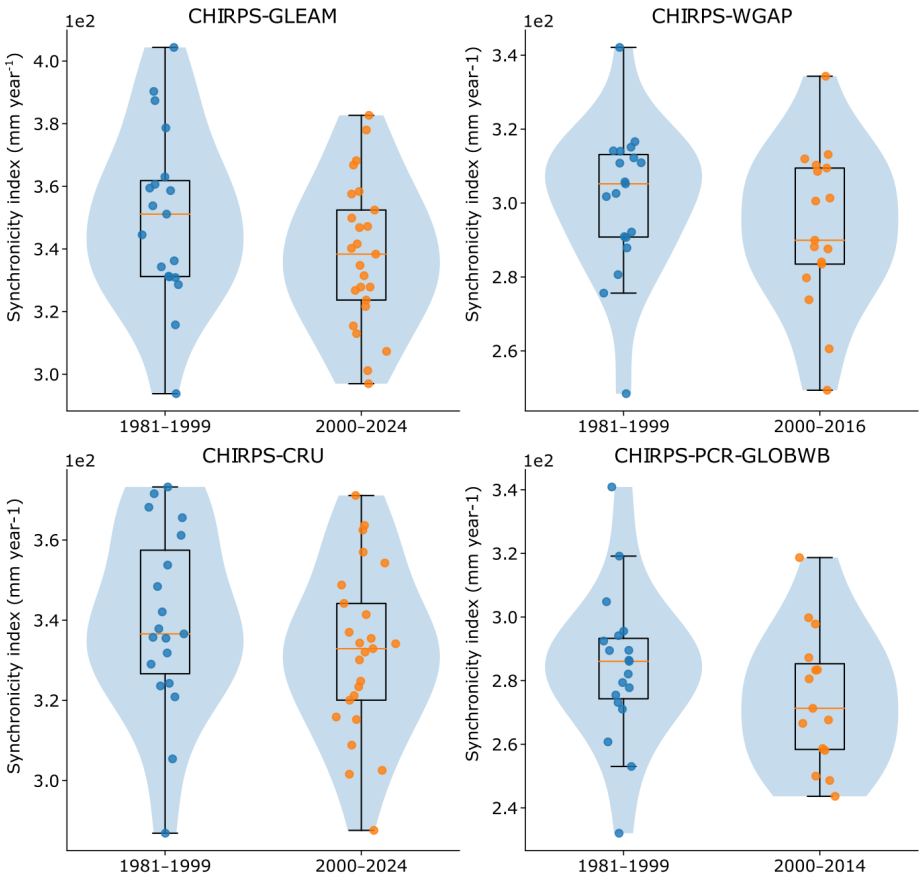


Figure 10. Violin and box plots of synchronicity index $\mathrm{mm\,year^{-1}}$ for the October–September hydrological year, derived from combining rainfall from CHIRPSv2 and four potential evaporation datasets (GLEAM v4.2a, WaterGAP3, CRU TSv4.09 and PCR-GLOBWB), for the periods before and after 2000.

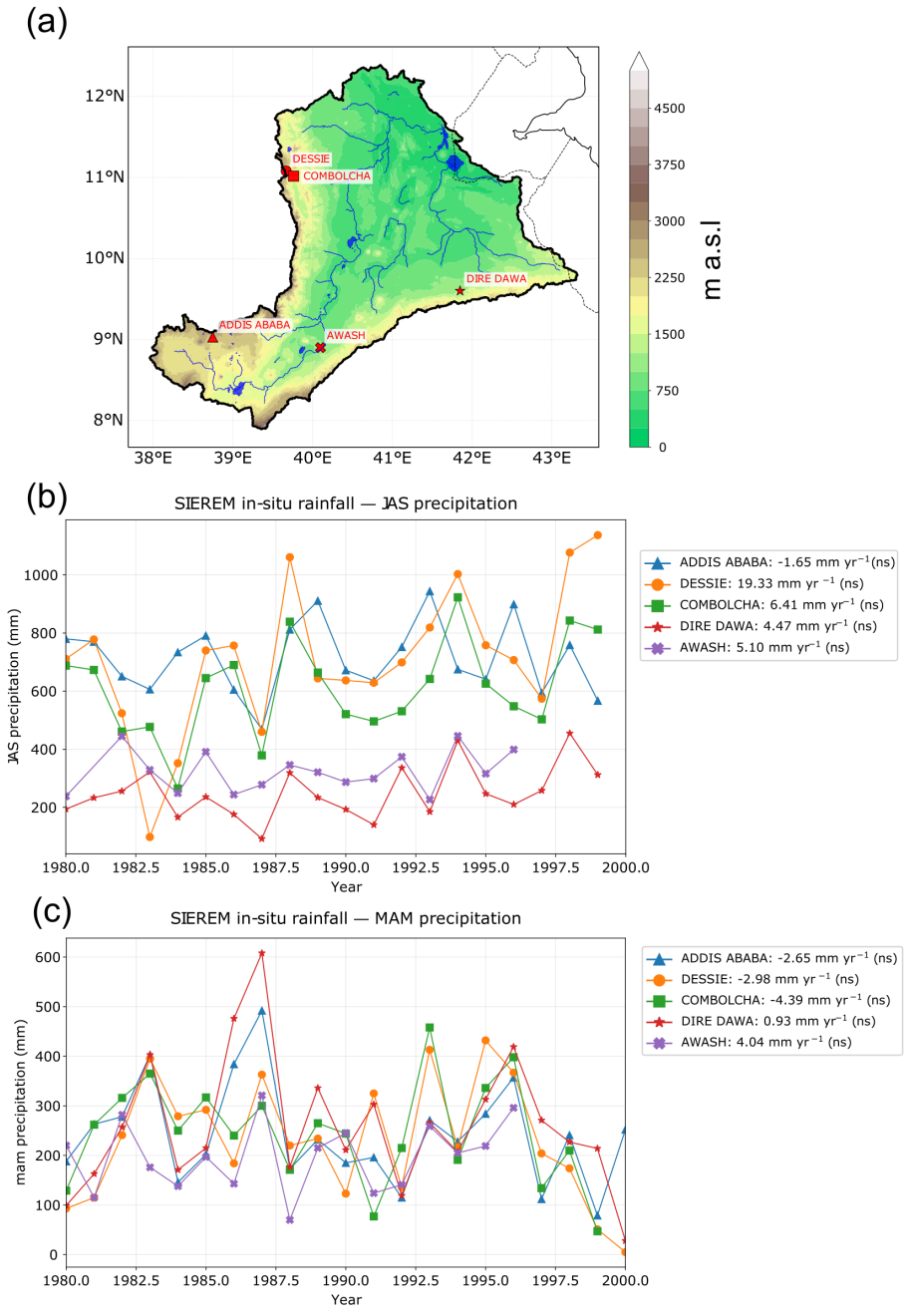


Figure 11. (a) Locations of SIEREM weather stations within the Awash Basin. Shading indicates elevation derived from the Shuttle Radar Topography Mission (SRTM) digital elevation model. Time series of monthly precipitation totals at selected stations with more than 10 years of available observations for (b) July–September (JAS) and (c) March–May (MAM). The estimated precipitation trends, calculated using Sen's slope estimator, are reported in the legend.

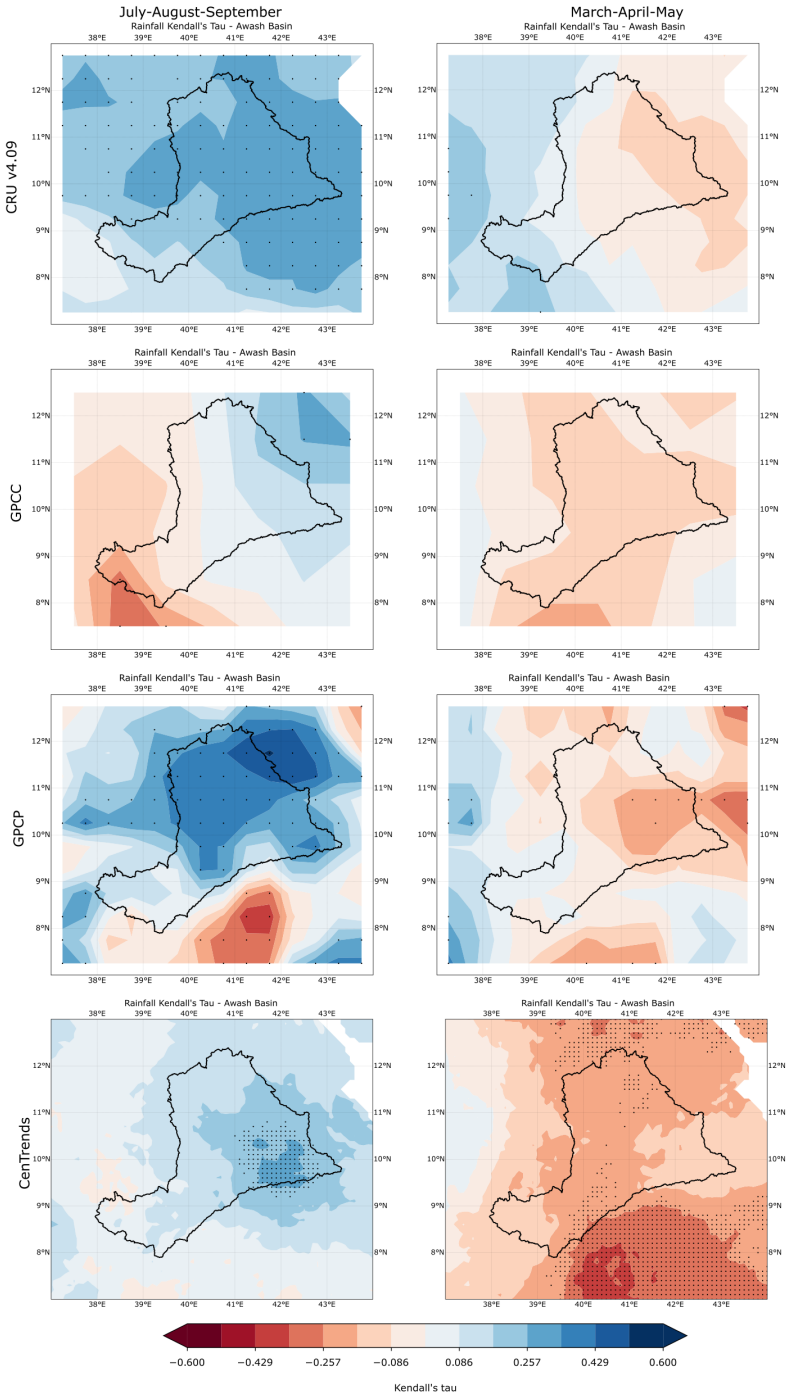


Figure 12. (Spatial trends in accumulatedrainfall over the Awash Basin during March–April–May (MAM) and July–August–September (JAS) over 1981–2024, respectively, based on CRUv4.09, GPCC, GPCP and CenTrends and expressed as Kendall's τ coefficients. Black dots indicate grid cells with statistically significant trends at the 95% confidencelevel ($p < 0.05$) using a Mann–Kendall test.