**--- Review 1 ---**

**Main comments**

The manuscript presents a novel framework that couples CWatM, which represents river routing, human water demand, and water resources management, with Earth system and land surface models through the OASIS3-MCT coupling interface. However, several aspects of the experimental design require further clarification, and the overall organization of the manuscript could be significantly improved. In addition, the Discussion should provide a more comprehensive assessment of the current limitations of the proposed framework and possible directions for future model development.

We sincerely thank the reviewer for their overall positive and thorough evaluation of the manuscript. We will carefully revise and improve the manuscript based on the comments and suggestions, and we are confident that it will better highlight the experimental design and potential limitations of our model. Below, you will find our detailed responses to the individual comments.

**Detailed comments:**

***Abstract***

1) *Line 5: Define OASIS3-MCT at it first occur.*
2) *Line 9: Define REMO at it first occur.*

We thank the reviewer for highlighting these points. OASIS3-MCT will be defined here as the "Ocean–Atmosphere–Sea–Ice–Soil Model Coupling Toolkit". REMO will be defined here as "REgional climate MOdel"

3) *Lines 11-12: Include quantitative performance metrics to support the reported improvements achieved by coupling, calibration, and bias correction.*

In the revised abstract, we will emphasise that the median KGE values across the selected set of GRDC basins increase by approximately 0.05 when comparing uncalibrated and calibrated runs, and by 0.1 after applying bias correction.

4) *Line 13: Define EURO-CORDEX at it first occur.*

EURO-CORDEX will be defined here as "the European branch of the COordinated Regional climate Downscaling EXperiment".

***1 Introduction***

5) *Lines 35-40: The Introduction places disproportionate emphasis on irrigation, whereas irrigation represents only one component of C-CWatM. The Introduction should instead emphasize the model's capability to represent human water demand and water resources management, which forms the primary motivation for the proposed framework.*

We agree with the reviewer that the current version of the introduction overemphasises irrigation. We will extend the paragraph like this:

"In addition to hydrological impacts, human water use and water resources management alter various land-surface processes, potentially modifying land-atmosphere interactions and local-to-regional hydroclimatic conditions. Therefore, their representation is essential to better understand human influences on water and energy cycles. Large-scale interventions ..."

6) Lines 61-63: The manuscript should better emphasize that the primary contribution of C-CWatM is the representation of realistic human water demand and water resources management. The methodology, results, and discussion should focus more on evaluating these capabilities rather than primarily on calibration and bias correction.

We fully agree with the reviewer that the representation of human water demand and water resources management is a central contribution of C-CWatM, and we will revise the manuscript to emphasise this more clearly in the Introduction, methodology, results, and discussion. However, calibration and bias correction are essential practical components for the effective application of C-CWatM in both offline and online coupled settings. As C-CWatM is a hydrological model, its performance is highly dependent on the quality of the land-surface forcing. While we keep these sections, we will revise the framing to make clear that they support the robust use of the coupling framework rather than presenting the main scientific contribution of the model.

7) *Line 66: According to the model description, the coupling is one-way rather than two-way, since information is transferred only from the land surface model to C-CWatM without feedback. The authors should use consistent terminology throughout the manuscript and avoiding expressions such as "fully coupled" or "two-way coupling" to prevent misleading readers.*

Thank you for this comment. We believe that we did not sufficiently distinguish between the coupling concept and the specific experiments presented in this study.

One of the main aims of developing C-CWatM is to enable a two-way exchange with the coupled climate model. In such a framework, the (regional) climate model provides the required forcing variables to C-CWatM, while C-CWatM returns water-management information, such as irrigation water amounts, to the climate model. Thus, C-CWatM is a model designed for coupled applications. However, we agree that the evaluations presented in this manuscript are based entirely on offline simulations. This distinction was not made sufficiently clear in the manuscript.

While the version currently archived on Zenodo (https://doi.org/10.5281/zenodo.19709771) only contains the functionality to receive the forcing fields, sending of the irrigation water amount has already been implemented and is currently being tested. Demonstrating and evaluating the complete two-way coupled framework additionally requires implementation of the corresponding coupling interface within the climate model and is beyond the scope of this model-description paper. These coupled experiments will be presented in a future publication.

To address your concerns and to avoid ambiguity, we will revise the manuscript to clearly distinguish between the current offline evaluation setup, which represents a one-way exchange of forcing fields, and the intended two-way coupled modelling framework for which C-CWatM has been developed. We will clarify this in the introduction, model concept, coupling interface and the corresponding results sections.

Please also refer to our response to points 12) and 18).

***2 Model concept***

8) *Lines 78-81: This section should focus primarily on introducing C-CWatM. The discussion of limitations of other models largely repeats the Introduction and could be shortened.*

The paragraph will be shortened to:

“Climate-CWatM (C-CWatM) is an integrated river-routing and water-resources model that uses land-surface variables as input. This design enables a direct linkage to state-of-the-art climate, Earth system, and land-surface modelling systems. Building on the experience of nearly two decades of global hydrological and water-resources modelling, C-CWatM simulates water availability, human water demand across various sectors, and water abstraction from different sources. The model can be operated offline, and online mode when integrated with climate or Earth system models through a coupling interface."

*9) Lines 81-83: Refer to Figure 1 when introducing the model framework.*

The figure reference will be implemented.

*10) Figure 1: Consider separating the online and offline configurations into two labeled subpanels (e.g., panels a and b) and refer to these subpanels consistently throughout the manuscript.*

We will restructure the figure as you suggested. See also the more detailed answer to point 20).

*11) Lines 85-86: Provide the governing equations or appropriate references for the major model components used in this research.*

In these lines, we refer to "land-surface processes such as runoff generation, soil water dynamics, and evapotranspiration", which are part of the parent model CWatM but were removed in C-CWatM to avoid duplication with the coupled land-surface model. The purpose of this paragraph is therefore only to explain the relationship between C-CWatM and its parent model. Readers interested in the governing equations and process representations of these components are referred to the detailed model description provided by Burek et al. (2020), which is cited here.

*12) Line 109: The coupling described appears to be one-way rather than two-way because variables are transferred from the land surface model to C-CWatM without feedback. revise the terminology accordingly.*

Please see our detailed answer to point 7 regarding the distinction between one-way and two-way coupling. To clarify this aspect, we will revise the text as follows:
"Coupling is implemented using the pyOASIS interface (Sect. 4.1), allowing C-CWatM (in Python) to exchange fields with any model featuring an OASIS coupling interface, independent of the programming language. In this study, the coupling interface is used to transfer forcing data to C-CWatM only, while the framework is designed to support future two-way coupling applications."

## *3 Overview of processes*

### *3.1 River routing, reservoirs and lakes*

*13) Lines 144-145: Provide the key equations (or references) describing the simplified representation of lakes and the treatment of open-water evaporation as an external forcing.*

In the original CWatM implementation, lake evaporation is estimated internally using the potential evaporation rate over open water surfaces. Potential evaporation is calculated within CWatM using the FAO Penman–Monteith method by default, with alternative methods also available (Burek et al., 2020). In C-CWatM, this calculation is omitted, and open-water evaporation is prescribed as an external forcing field provided by the coupled climate model. Consequently, no additional parameterization or equations are required within C-CWatM, as the evaporation flux is computed by the land surface scheme of the coupled climate model. Other aspects of the lake parametrisation have not been modified. We will clarify this as follows:
"In the original CWatM implementation, lake evaporation is estimated using the potential evaporation rate over water surfaces (Burek et al. 2020). In C-CWatM, this calculation is omitted, and open-water evaporation is prescribed as an external forcing field provided by the coupled climate model."

### *3.3 Human water use*

#### *3.3.1 Irrigation water demand*

*14) Line 174: Provide a reference supporting the assumption of maintaining a 50 mm water depth for paddy irrigation.*

The assumption of maintaining a 50 mm ponding depth for paddy irrigation inherited from the parent model CWatM. The paddy irrigation scheme is described in detail by Wada et al. (2014), where this assumption is adopted based on studies by Wisser et al. (2008,2010). We will modify the corresponding sentence to:

“In the original implementation of CWatM, paddy irrigation is represented by maintaining a surface water depth of 50 mm (Wada et al., 2014), with water demand calculated as the difference between current and target water levels.”

Wada, Y., Wisser, D., and Bierkens, M. F. P.: Global modeling of withdrawal, allocation and consumptive use of surface water and groundwater resources, Earth Syst. Dynam., 5, 15–40, https://doi.org/10.5194/esd-5-15-2014, 2014.

Wisser, D., Frolking, S., Douglas, E. M., Fekete, B. M., Vörösmarty, C. J., and Schumann, A. H.: Global irrigation water demand: Variability and uncertainties arising from agricultural and climate data sets, Geophys. Res. Lett., 35, L24408, https://doi.org/10.1029/2008GL035296, 2008.

Wisser, D., Fekete, B. M., Vörösmarty, C. J., and Schumann, A. H.: Reconstructing 20th century global hydrography: a contribution to the Global Terrestrial Network- Hydrology (GTN-H), Hydrol. Earth Syst. Sci., 14, 1–24, https://doi.org/10.5194/hess-14-1-2010, 2010.

*15) Lines 177-179: Provide references supporting the simplified irrigation representation adopted in the model.*

In the revised manuscript, we will clarify that simplified paddy irrigation representation in C-CWatM is a pragmatic approximation, as the surface-water depth used in the original CWatM paddy scheme is generally not available from coupled land-surface or climate models. We therefore approximate the absence of water stress in paddy-irrigated fields by maintaining root zone soil moisture the highest attainable target (field capacity; represented by an alpha-depletion factor of α = 1). In contrast, non-paddy irrigation uses lower depletion thresholds. We will add the reference to the original CWatM irrigation scheme and explicitly state that the C-CWatM paddy formulation is a simplified approximation to enable coupling. We will rewrite the paragraph as follows:

"... cannot be provided. We therefore adopted a simplified approach to represent the flooded conditions of paddy irrigation in C-CWatM. In this approximation, non-water-stressed conditions are maintained for the crop, which results in a target soil water content at field capacity and an alpha-depletion factor of α = 1. This formulation should be viewed as a practical approximation rather than an explicit representation of paddy irrigation.”

*3.3.3 Water abstraction by source*

*16) Lines 194-197: This subsection represents one of the key contributions of C-CWatM. The ability to allocate water demand from different sectors (e.g., irrigation, domestic, industrial, and livestock) among multiple water sources (surface water and groundwater) is one of the major strengths of the model. The authors should expand this description and provide the governing equations or appropriate references for the water allocation scheme.*

The water allocation schemes are the same as in CWatM (Burek et al. 2020), which are based on the work of Hanasaki et al. (2018) and the implementation in the hydrological model PCR-GLOBWB 2 (Sutanudjaja et al. 2018), which are based on de Graaf et al. (2014) and Wada et al. (2014). We will expand the description to clarify the abstraction hierarchy and add the mentioned references as follows:

“Water withdrawal from available sources follows a hierarchical abstraction scheme adapted from CWatM. In this dynamical water allocation approach, abstractions are based on the availability of individual water sources (de Graaf et al. 2014, Wada et al. 2014) rather than on prescribed fractions. By default, water is drawn in the following sequence: surface water is utilized first, starting with channels, followed by big lakes and reservoirs, and finally small lakes. If surface water alone cannot meet the demand, additional water is drawn from renewable groundwater.

If local supply is insufficient, the model expands the search domain and draws water from a buffer zone around lakes and reservoirs, followed by surface water and groundwater in neighbouring cells within a predefined abstraction radius. This representation accounts for the fact that water can be transported through water-supply infrastructure beyond the local grid cell.

If the demand is still not met, users can configure whether the remaining demand should persist as unmet demand or be fulfilled by non-renewable ground water, by setting the *limitAbstraction* parameter in the settings file. The use of non-renewable groundwater corresponds to unsustainable groundwater use, where abstractions are not compensated by groundwater recharge. The non-renewable groundwater contains an unlimited amount of water (Hanasaki et al. 2018)."

de Graaf, I. E. M., van Beek, L. P. H., Wada, Y., and Bierkens, M. F. P.: Dynamic attribution of global water demand to surface water and groundwater resources: Effects of abstractions and return flows on river discharges, Adv. Water Resour., 64, 21–33, https://doi.org/10.1016/j.advwatres.2013.12.002, 2014.

Hanasaki, N., Yoshikawa, S., Pokhrel, Y., and Kanae, S.: A global hydrological simulation to specify the sources of water used by humans, Hydrol. Earth Syst. Sci., 22, 789–817, https://doi.org/10.5194/hess-22-789-2018, 2018.

Sutanudjaja, E. H., van Beek, R., Wanders, N., Wada, Y., Bosmans, J. H. C., Drost, N., van der Ent, R. J., de Graaf, I. E. M., Hoch, J. M., de Jong, K., Karssenberg, D., López López, P., Peßenteiner, S., Schmitz, O., Straatsma, M. W., Vannametee, E., Wisser, D., and Bierkens, M. F. P.: PCR-GLOBWB 2: a 5 arcmin global hydrological and water resources model, Geosci. Model Dev., 11, 2429–2453, https://doi.org/10.5194/gmd-11-2429-2018, 2018.

Wada, Y., Wisser, D., and Bierkens, M. F. P.: Global modeling of withdrawal, allocation and consumptive use of surface water and groundwater resources, Earth Syst. Dynam., 5, 15–40, https://doi.org/10.5194/esd-5-15-2014, 2014.

*17) Line 203: Clarify how the environmental flow threshold is determined, including the data source, parameter values, and corresponding references.*

C-CWatM retains the implementation in the parent model CWatM, where the environmental flow requirements are not determined by a threshold or a fixed percentage of streamflow. Instead, the model provides the ability to use pre-computed maps of environmental need, which are read as model input (Burek et al. 2020, CWatM documentation). C-CWatM does not provide environmental flow maps itself, and this functionality is not used in the simulations presented in this study. We will clarify this as follows:

"In addition to human water demand, C-CWatM can account for environmental flow requirements provided as externally pre-computed maps. These user-specified fields represent the minimum flow required to sustain ecological functions in rivers, reservoirs, and lakes."

Community Water Model CWatM Manual, Version 1 for CWatM 1.05, https://pure.iiasa.ac.at/id/eprint/16802/1/CWatM_Manual_v1.pdf (last access 15 July 2026)

## *4 Model enhancements and optional features*

### *4.1 The OASIS3-MCT coupling interface*

*18) Lines 208-209: Based on the model description, the coupling appears to be one-way, avoid the term "fully coupled" and clarify that OASIS3-MCT transfers information from the land surface model to C-CWatM rather than performing two-way data exchange.*

Please see our detailed anwer regarding the distinction between one-way and two-way coupling in point 7). To clarify this aspect, we will revise the text as follows:

“An efficient way to **exchange data** between C-CWatM and (regional) climate models is to use an established coupling software that handles the synchronized data exchange and regridding of the coupling field.”

*19) Line 227: Clarify the computational benchmark used to quantify the reported 50% increase in runtime, including the computing platform and the baseline configuration.*

The benchmark was performed on the Levante HPC system using one compute node on the compute partition. We compared the wall-clock time of the original C-CWatM configuration using preprocessed forcing fields with the configuration reading and regridding the REMO forcing data online through the OASIS interface. The reported ~50% increase is based on the elapsed wall-clock time reported by the job scheduler for five repeated model runs. We will clarify this as follows:

“In our benchmark tests on the Levante HPC system, reading and regridding the REMO forcing data through the OASIS interface increased the model runtime by approximately 50% compared to the configuration using preprocessed forcing fields. The additional computational cost of this approach remains acceptable and eliminates the offline preprocessing step previously required for regridding.”

*20) Figure 1: The role of the OASIS3-MCT coupling interface should be highlighted more clearly, as it represents one of the key methodological contributions of this study. If any part of the figure is adapted from existing CWatM documentation or other published sources, provide the appropriate citation. In addition, the current figure contains a large amount of secondary graphical information that is not essential to understanding the coupled framework. A simplified conceptual diagram focusing on the interaction between the land surface model, OASIS3-MCT, and C-CWatM would improve readability.*

Thank you for this suggestion. We will revise the figure according to your suggestions by distinguishing between online and offline configurations and by highlighting the coupling interface.

The central part of the figure (schematic 3D image of a socio-hydrological model) is taken from CWatM. As the version displayed in the CWatM model desciption paper (Burek et al. 2020, Fig. 1) contains process labels, we used the plain version provided in the CWatM GitHub repository. The already existing citation in the figure caption will be adapted to:

“This center schematic of this figure is based on material from the CWatM GitHub repository (https://github.com/iiasa/CWatM, last access: 24.04.2026) modified and used under the GNU General Public License v3.0.”

*4.2 Optional model calibration*

*21) Line 247 and Table 1: Provide the justification or reference for selecting alpha depletion is 0.7 as the newly introduced calibration parameter.*

The concept of an alpha depletion parameter, which defines the soil water depletion level that triggers irrigation, is inherited from the parent model CWatM. This concept is a simplified implementation of the approach based on soil moisture depletion described in Wada et al. (2014). While CWatM uses a default value of 0.7, we consider the alpha depletion parameter to vary for different regions, crops and irrigation practices and thus recommend to use it as a calibration parameter. We will rephrase the paragraph to:

“We extended the list by the ’alpha depletion’ parameter, which defines the upper soil moisture threshold up to which farmers irrigate. This parameter is likely of importance for studies focused on irrigation and could also be set to a fixed value based on a priori knowledge. While a default value of 0.7 is adopted from the parent model CWatM, the depletion threshold may vary for different regions, crops, and irrigation practices. Therefore, we included it as a calibration parameter."

Wada, Y., Wisser, D., and Bierkens, M. F. P.: Global modeling of withdrawal, allocation and consumptive use of surface water and groundwater resources, Earth Syst. Dynam., 5, 15–40, https://doi.org/10.5194/esd-5-15-2014, 2014.

*4.3 Bias correction*

*22) Lines 257-263: The discussion of bias correction and its limitations is unnecessarily lengthy. This section could be simplified into one or two concise paragraphs while focusing on the bias correction method actually implemented in this research.*

We agree that the discussion can be condensed. However, we believe that a brief discussion of the motivation for bias correction is useful in a model-description paper. We also specifically address the challenges arising in online-coupled simulations, where conventional preprocessing-based bias correction is not feasible. We will shorten the paragraph to:

"Bias correction is often used in climate and impact modelling to adjust systematic errors in climate model outputs by considering observations or observation-based data. Hydrological models are particularly sensitive to forcing biases, as errors in water-cycle variables can propagate through the model and affect estimates of streamflow, soil moisture, and water management. Consequently, impact studies commonly apply methods such as quantile mapping to adjust forcing data prior to hydrological simulations."

23) Line 275: The real-time bias correction is estimated from reference CWatM simulations rather than observational datasets. This experimental design makes it difficult to distinguish improvements arising from the coupled model itself from those introduced by the bias correction procedure. Please discuss this limitation more thoroughly.

We fully agree and will clarify this issue in the revised version of the manuscript. Please also refer to the remaining comments on bias correction. Since the runoff bias correction is based on CWatM reference simulations, any improvement observed after bias correction cannot be interpreted independently. We will revise the manuscript to highlight the experiments as a proof of concept for C-CWatM's ability to implement bias correction and assess runoff biases related to the forcing.

*4.4 Reduced forcing requirements*

*24) Line 280: Define CMIP and CORDEX at it first occur.*

CMIP will be defined as "Coupled Model Intercomparison Project". CORDEX will have been already defined in the abstract as "COordinated Regional climate Downscaling EXperiment".

*25) Line 281: Refer to the corresponding panel in Figure 1 when introducing the offline configuration.*

The figure reference will be implemented.

*5 Data*

*5.1 Input map*

*26) Lines 301-302: Identify the dataset used for field capacity and wilting point, including its spatial resolution. If this dataset is not already included in Table B1, it should be added accordingly.*

We will add the two datasets to Table B1 and update the description as follows:

"For offline simulations, we supply precomputed field capacity and wilting point datasets generated using the same method as in CWatM, based on HSWD parameters and the Mualem- and van

Genuchten model. These datasets correspond to the original CWatM soil water storage parameters and are provided in spatial resolutions of 0.5° and 5' (Table B1)."

*27) Figure 2: Add a north arrow and scale bar.*

This will be done.

*5.2 Forcing data*

*28) Table B-1: Clarify how the 3′ DEM was resampled to match the 5′ model grid.*

We did not perform any DEM resampling as part of this study. The model uses the preprocessed 5′ static elevation standard deviation dataset provided with CWatM, which is derived from SRTM (Jarvis et al., 2008) with 3'' resolution and Hydro1k (USGS, 2002) with 1 km resolution. For clarification, we will add the following footnote to Table B1:
"The model uses the 5′ elevation dataset provided with CWatM (https://drive.google.com/drive/folders/1HqcBj5fD6DHJpOe-t_6JHFMKFtubghZf, last access: 24 April 2026); no additional resampling was performed in this study."

29) Lines 307-308: Describe how the 0.11° EURO-CORDEX forcing was transformed to the 5′ C-CWatM grid.

All data have been transformed from 0.11 degrees to 5 arcmin using bilinear remapping. We will provide this information in the revised manuscript.

30) Line 323-325: Provide the reference describing the conversion from relative soil saturation (SM relsat) to absolute soil moisture.

The percentile-based conversion to relative soil saturation is a practical approximation introduced in this study for reduced-forcing applications, rather than a method adopted from a specific reference. It is used because consistent model-specific soil hydraulic parameters and field-capacity information are often unavailable for EURO-CORDEX simulations. We will change the last sentence of the paragraph to:

"This method is a practical approximation that enables us to convert the variable mrso into relative saturation for each model. "

*5.3 CWatM model run for comparison*

*31) Line 328: Clarify the spatial resolution of the MSWX forcing dataset.*

The resolution is 0.1 degrees. We will add this information as follows:
"The reference simulation is driven by MSWX meteorological forcing (Beck et al., 2022), which provides globally consistent, observation-based estimates of precipitation, temperature, and other meteorological quantities every 3 hours at a high spatial resolution of 0.1°."

***6 Test application***

*32) Lines 336-337: The description of the CDO bilinear remapping applied to the REMO outputs is part of the forcing data preprocessing and would be more appropriately presented in Sect. 5.2 (Forcing data) rather than in Sect. 6 (Test application). The relevant parts of Sect. 6 should be integrated into Sect. 5.2.*

We will move the corresponding sentences to Section 5.2.

*6.1 River discharge and water demand*

33) Lines 342-343 and Figure 3: The manuscript states that the model reproduces the European river discharge patterns well. Figure 3 only presents the spatial distribution of simulated discharge.

The authors should add the spatial distribution of the reference discharge together with a difference map (simulated minus reference discharge) to better support this conclusion.

34) Line 350 and Figure 3: If no observational dataset is available for evaluating water demand, the authors should move these results to the Appendix or Supplementary Material rather than presenting them in the main figure. The main figure should focus on results that directly support the principal conclusions of the manuscript.

We thank the reviewer for these suggestions, which we will implement in the revised version of the manuscript. The revised Figure 3 will feature a comparison between simulated and reference discharge. Water demand maps will be presented as a separate figure in the Appendix.

35) Line 359: The experimental design requires further clarification. The coupled C-CWatM simulations are driven by runoff, evaporation and soil moisture simulated by REMO2020, whereas the standalone CWatM reference simulations are driven by MSWX meteorological forcing. Consequently, the differences may arise from both the forcing datasets and the coupled modeling framework, making it difficult to attribute the performance improvements solely to model coupling.

The authors should clarify the rationale for this experimental design and discuss the influence of the different forcing datasets on the comparison. Ideally, the baseline CWatM simulations should be driven by forcing data that are consistent with those used in the coupled simulations to better isolate the effect of the coupled modeling framework.

We agree that the comparison between C-CWatM driven by REMO and the CWatM reference simulation driven by MSWX forcing does not isolate the effects of the coupled modelling framework alone. We will clarify this more prominently in the revised manuscript. It is important to note that the goal of this comparison is not to demonstrate that coupling improves performance over CWatM simulations. CWatM is a widely used and validated large-scale hydrological model. By comparing our C-CWatM simulations to CWatM, we aim to evaluate whether C-CWatM driven by land-surface output from a climate model can reliably reproduce spatial discharge patterns. We evaluate the spatial pattern relative to the "reference" CWatM-MSWX simulation, which provides a spatially continuous hydrological benchmark, highlighting the potential capabilities of the modelling system when using dedicated forcing datasets. We do not consider the CWatM-MSWX simulations as the observational ground truth. Independent validation is provided through comparison against GRDC discharge observations. We will revise the manuscript accordingly.

Additionally, we will add CWatM simulations driven by ERA5-based forcing to better illustrate the role of forcing uncertainty. See also our response to point 38.

*6.2.1 Validation against observed discharge*

36) Lines 376-377 and Figure 4: To better demonstrate the added value of the coupled model, the authors should include the performance of the stand-alone CWatM against GRDC observations in Fig. 4e, together with a spatial distribution of KGE for the standalone CWatM. Such a comparison would provide a more direct assessment of whether coupling improves model performance relative to the standalone configuration.

We thank the reviewer for that suggestion, which will help to clarify the intention of our comparison between C-CWatM simulations driven by REMO and CWatM simulations driven by MSWX forcing. Additionally, we will include the performance results of CWatM simulations driven by ERA5 to better illustrate the impact of forcing uncertainties.

37) Lines 390-393: Further discuss why the largest discrepancies occur over southern and central Europe, whereas the strongest positive biases are mainly found over northern and eastern Europe.

We will add a short explanation in the revised version of the manuscript highlighting that these patterns likely reflect a combination of regional biases in REMO land-surface variables and hydrological process representation in C-CWatM. We will add the following text:

“Regional biases likely correspond to a combination of systematic biases in the REMO land-surface forcing and regional differences in hydrological process representation in C-CWatM. Positive biases in northern, eastern, and mountainous regions may partly relate to biases in runoff generation, snow accumulation and melt, whereas negative biases in western and southern Europe may be influenced by biases in precipitation, evapotranspiration, and soil-moisture dynamics.”

*6.2.2 Validation against reference simulations*

38) Lines 416-417: Improved agreement with the reference model does not necessarily imply improved agreement with observations. Since the standalone CWatM runoff is used as the reference for bias correction, the improved consistency between the two models is an expected outcome. These results should be interpreted with greater caution.

We fully agree and will interpret the results with greater caution. Since CWatM reference runoff is used for bias correction, improved agreement with CWatM reference discharge is partly expected and should not be interpreted as independent validation (which was never our intention). We will revise the text accordingly, highlighting that our analysis aims to highlight the general capability of C-CWatM to represent spatial discharge patterns. Independent validation is provided by GRDC discharge observations. In addition, CWatM simulations driven by ERA5 will help illustrate the role of forcing uncertainty. Please also refer to our response to point 35.

39) Lines 419-420: Because bias correction artificially increases the consistency between the two model configurations, the conclusion that bias correction is more important than calibration is not yet sufficiently supported. The manuscript does not demonstrate whether a similar conclusion would still hold if bias correction were performed using independent observations (e.g., GRDC). If the primary objective of the study is to demonstrate the added value of C-CWatM, the discussion should focus more on the comparison between the coupled and standalone model configurations (e.g., noC-noBC versus C-noBC), rather than primarily emphasizing the improvements introduced by bias correction.

40) Figure 7: The current figure mainly highlights the benefit of bias correction. The authors should redesign this figure to emphasize the advantages of the coupled model itself, for example by comparing C-CWatM simulations driven by REMO runoff directly against observations and against the standalone CWatM simulations.

We agree and will interpret all results related to the bias correction with greater caution. Since biases are corrected based on CWatM reference runoff, improved agreement with the CWatM reference simulation is expected and should not be interpreted as independent validation. We will provide a more nuanced interpretation in the revised version of the manuscript to avoid creating the impression that bias correction is generally more important than calibration. We will revise the text accordingly and frame the experimental setup more as a proof of concept, aiming to demonstrate the technical capability and potential usefulness of runoff bias correction.

We will further clarify that GRDC provides observed discharge at gauging stations. It is not possible to use GRDC observations as spatially distributed reference runoff for the implemented runoff bias correction at 5 arcmin across Europe. GRDC is used for independent validation at single basins, while the CWatM reference simulation is used as a spatially continuous hydrological benchmark, highlighting the potential capabilities of the modelling system when using dedicated forcing datasets (see our previous responses). We will revise Figure 7 and the associated discussion, clarifying that the reduction of REMO runoff biases relative to the CWatM runoff benchmark is shown (which is expected).

### *6.3 Reduced forcing considering EURO-CORDEX simulations*

41) Figure 8: The authors should compare the C-CWatM simulations forced by the 13 EURO-CORDEX model outputs directly against GRDC observations rather than against the standalone CWatM reference simulations. Such a comparison would provide a more direct assessment of the simulation ability of the coupled model under different climate forcing datasets.

We will provide a GRDC-based evaluation of the EURO-CORDEX-driven simulations in the Appendix of the revised version. The resulting KGE values are generally low. This can be expected for raw, non-bias-corrected EURO-CORDEX forcing combined with the simplified reduced-forcing assumptions. In comparison, Figure 8 evaluates how well C-CWatM represents spatial discharge patterns across the European river network when driven by readily available climate-model output. The CWatM reference simulation is used as a spatially continuous hydrological benchmark, not as observational ground truth. We will revise the text as follows:

“The comparison in Fig. 8 provides a spatial analysis rather than an independent observational validation. We use the CWatM reference simulation as a spatially continuous hydrological benchmark to assess the capability of reduced-forcing C-CWatM simulations to represent large-scale discharge patterns. An additional comparison against GRDC observations is provided in Appendix X. The relatively low KGE values obtained for the raw EURO-CORDEX-driven simulations highlight the importance of climate-forcing biases and the limitations of the simplified reduced-forcing assumptions.”

## *7 Discussion*

42) Lines 455-459: The authors state that the limited availability of GRDC observations prevents their use for bias correction; however, this alone does not fully justify the chosen approach. Bias correction could still be evaluated for basins with sufficient observations, followed by a discussion of the corresponding results. In addition, the manuscript should discuss the implications of using model-simulated runoff, rather than observational data, as the reference for bias correction, particularly with respect to the interpretation of the coupled model performance.

We agree that the interpretation of using CWatM-simulated runoff for bias-correction requires a more thorough discussion. We will revise the manuscript accordingly (please also refer to our previous responses). However, we would like to clarify that GRDC provides observed river discharge at gauging stations. CWatM and C-CWatM require spatially distributed runoff fields, and GRDC observations cannot be used directly as a runoff reference, even for individual basins. We therefore use CWatM reference runoff as a practical proof-of-concept for the runoff bias-correction approach, while clearly stating that our analysis does not permit independent evaluation of bias-corrected C-CWatM simulations. We will revise the text as follows:

“CWatM reference runoff is used here as a practical, spatially distributed benchmark to demonstrate the feasibility of runoff bias correction, as no observations-based, high-resolution, gridded runoff dataset is available across Europe. The resulting assessment of bias-corrected C-CWatM simulations relative to the CWatM reference simulation should not be interpreted as independent observational validation.”

43) Line 465: The manuscript should provide further discussion of the potential influence of using different meteorological forcing datasets for the standalone CWatM (MSWX) and the coupled C-CWatM (REMO2020). Since differences in model performance may arise from both the forcing data and the coupled modeling framework, the contribution of each source of uncertainty should be discussed more explicitly.

In addition, given that REMO2020 is driven by ERA5 boundary conditions, please justify why the standalone CWatM simulations were not performed using ERA5-based forcing to provide a more consistent baseline for evaluating the coupled modeling framework.

We will provide a more thorough discussion of forcing uncertainties in the revised version of the manuscript and add CWatM simulations driven by ERA5 forcing as an additional comparison to help assess the potential influence of the forcing dataset. However, it is important to note that CWatM requires meteorological and C-CWatM land-surface forcing. REMO2020 is only constrained by ERA5 at the boundaries and generates independent regional land-surface fields within the domain. We use MSWX as the primary CWatM forcing dataset because it is better suited for hydrological applications and provides bias-adjusted meteorological forcing. We will therefore carefully interpret the ERA5 comparison as an additional sensitivity experiment.

44) Lines 459-460: The manuscript reports predominantly positive discharge biases over northern, eastern and mountainous regions, whereas negative biases mainly occur over western and southern Europe. The physical mechanisms responsible for these regional differences deserve further discussion.

We will add a further discussion in the revised version of the manuscript highlighting that these patterns likely reflect a combination of regional biases in REMO land-surface variables and hydrological process representation in C-CWatM (see also our response to point 37).

45) The Discussion would benefit from a more comprehensive assessment of the current limitations of C-CWatM and possible future improvements. Besides groundwater and irrigation processes, the authors should discuss how the coupling scheme, the choice of forcing datasets, and the bias-correction framework may contribute to the remaining model biases. A broader discussion of potential improvements would provide a more balanced assessment of the strengths and limitations of the coupled modeling framework.

We agree with the reviewer that the Discussion should provide a broader and more balanced assessment of the current limitations of C-CWatM, including groundwater and irrigation processes, coupling issues, the use of different forcing datasets, and bias-correction. We will also outline possible future improvements in more detail.