

<RESPONSE TO REFEREE #2>

- **Title: CAMELS-KR: Catchment attributes, meteorology, and streamflow for large-sample hydrology in South Korea**
- **Authors: Songyun Lee, Guchang Jung, and Kuk-Hyun Ahn**

((Acknowledgment))

We thank Reviewer #2 for their thoughtful and constructive comments. We appreciate the reviewer's careful evaluation and have addressed each point below. Please note that all line numbers referenced in this response correspond to the non-tracked-changes version of the revised manuscript.

Specific Comment

((Comment #1)) Title, specifically the phrase "reconstructed streamflow". As with many similar datasets, CAMELS-KR contains modelled streamflow, and I guess this is what is meant by "reconstructed". However, a key point of any CAMELS dataset is making the observed streamflow available (since this is not usually available via other means). Using the phrase "reconstructed streamflow" gives the impression that the dataset doesn't contain observed streamflow, which is misleading (because it isn't true!).

((Response #1)) We agree that the phrase “reconstructed streamflow” in the title could give the misleading impression that CAMELS-KR does not include observed streamflow. As the reviewer pointed out, making observed streamflow available is a key contribution of CAMELS datasets. We have therefore removed “reconstructed” from the title, replacing “reconstructed streamflow” with “streamflow” to better reflect the dataset's contents.

((Comment #2)) Level of nestedness. A key reason the authors have been able to include such a large set of catchments from such a small space is that much of the dataset is composed of sets of gauges moving downstream along the same river system. All CAMELS datasets have this issue to some extent, so it is not a serious issue. However, it is a problem that it is not emphasised in the manuscript, nor the issues discussed (e.g. it means the timeseries of these

catchments have a lot in common and are not independent sources of information). I suggest the authors: (i) make this clearer, including in the abstract; (ii) quantify it via attributes in the dataset - see e.g. the CAMELS-Australia or CAMELS-Peru datasets for examples; (iii) include a figure that summarises the nestedness statistics; and (iv) devote a paragraph to discuss the implications.

((Response #2)) The reviewer is correct that the nested structure of the CAMELS-KR catchments was not sufficiently emphasised in the original manuscript. Multiple gauges are located along the same river systems, and the resulting spatial overlap means that the 282 catchments should not be regarded as independent sources of hydrological information. We now explicitly highlight the presence of nested catchments in the abstract (line 13) and clarify in the main text that the number of catchments partly reflects the inclusion of multiple gauging locations within the same river networks (lines 164-167).

To describe these nesting relationships, we have added two attributes, `is_headwater` and `upstream_basin_ids`, to the `CAMELS_KR_location_attributes` file. The `is_headwater` attribute indicates whether a catchment has no upstream catchment within the CAMELS-KR network, while `upstream_basin_ids` lists the identifiers of the CAMELS-KR catchments located upstream of each catchment. These attributes allow users to identify nesting relationships and determine the number of upstream catchments represented in the dataset. Their definitions have been added to Section 4.1, Table 2, and the dataset documentation. We also clarify that `is_headwater` refers specifically to the CAMELS-KR network and does not necessarily identify a physical headwater catchment.

((Comment #3)) There is clearly a lot more water-level data than discharge data. I have two suggestions, listed below:

(i) Figure 3 is fantastic and very useful, but it doesn't help the reader to understand the catchment-by-catchment availability of discharge vs water levels. I suggest to add a panel (d) here based on the data in panel (c) but plotting "years of available water level data" on one axis and "years of available discharge data" on the other axis, and each scatter point is a catchment.

(ii) Assuming many catchments where water level data extends backwards more than discharge,

surely an opportunity might be to develop a method to estimate discharge for these earlier periods, based on water level. I am not suggesting the authors actually do this here, but if the authors agree it is an opportunity, then perhaps discuss it in Section 7 and note any potential issues that would need to be overcome or that would increase uncertainty in the outcomes.

((Response #3)) Based on the reviewer's recommendation, we have added panel (d) to Figure 3 to compare the number of years with available water-level and discharge observations for each catchment. We also agree that historical water-level observations could provide an opportunity to extend streamflow time series where they predate the available discharge records. We have revised the final paragraph of Section 7 to discuss this potential extension using site-specific stage–discharge relationships. We also highlight potential limitations, including changes in channel geometry and water-level reference datums.

((Comment #4)) Catchment delineation method is not described, or at least, I can't find it. Line 248 claims it is described in Section 2, but it doesn't seem to be? On a related topic, catchment area calculation is very important for streamflow conversions to mm/d but is not given sufficient detail. Specifically, a spatial projection would have been needed to express area in square kilometres, but this does not seem to be stated. I suggest this be rectified in the text and also specified in Table 2 "basin_area" row.

((Response #4)) We admit that the original manuscript did not adequately describe the catchment delineation procedure or the calculation of catchment area. We have expanded Section 2.2 (Lines 126–137) to describe the Python-based delineation workflow, including DEM conditioning, generation of flow-direction and flow-accumulation grids, adjustment of gauge locations to the DEM-derived drainage network, and conversion of the delineated catchments into vector boundaries.

In addition, we now explicitly describe how catchment areas were calculated using [area calculation method] in [projected coordinate reference system; EPSG code], with the resulting areas expressed in square kilometres. The description of basin_area in Table 2 has also been revised to specify the coordinate reference system.

((Comment #5)) Data quality. There seems to be several indications of data quality issues, possibly associated with errors and/or human activities. For example, the LSTM and rainfall-runoff results exhibit quite low performance (see separate comment below) while Figure 7 seems to indicate that an unexpectedly large proportion of catchments do not meet basic water balance constraints, with streamflow seemingly exceeding precipitation or, alternatively, implied AET ($P-Q$) exceeding PET. Again, this is not a problem per se, since data quality is difficult to assure across large datasets. However, it needs to be discussed in more detail. Why do the catchments not meet water balance constraints? In particular, how much of it is due to human activities shifting water from place to place, how much of it is due to data errors (etc.)?

((Response #5)) We agree that departures from basic water-balance constraints require a more explicit discussion of both human influences and potential data uncertainties. We have expanded the discussion accompanying Figure 7b to identify 24 catchments that fall beyond the theoretical energy limit and 16 catchments with runoff ratios greater than unity ($Q/P > 1$). However, the available information does not allow us to quantify the relative contributions of human activities and data errors to the observed departures. The spatial association with urban areas suggests possible anthropogenic influences but does not establish their causes. Separating these contributions would require catchment-specific information on water transfers, withdrawals, return flows, and storage changes, together with further evaluation of the precipitation and discharge records. We therefore present these explanations as possible contributing factors rather than confirmed causes and caution users when interpreting the affected catchments.

((Minor comment #6)) A few parts of the text unnecessarily repeat lists, at the cost of brevity. Examples include lines 310-312, which should be combined into lines 308-309. Similarly, the information on catchment selection criteria is presented once as a list of definitions (starting L134), then again as a list of reasons (starting L142)--it is recommended that this would be better suited to a table (Criterion || Description || Reason || Number of catchments that meet this rule). On a related note, at Line 135: the provision of "n = xxx stations" here is very useful. I suggest to clarify that decrease in numbers from one rule to the other is the result of application of the new rule to the set of catchments proceeding from the previous rules. That is, the text "[stations must] provide at least 10 years of discharge data (314 stations)" should not be read

as "there are 314 stations in South Korea with >10 years of data" but instead that the 314 stations out of the 850 identified in the previous step qualify.

((Response #6)) Following the reviewer's recommendation, we have merged the repeated lists of land-cover classes and their corresponding attributes into a single, concise description. We have also consolidated the catchment selection criteria into Table 1, which presents each criterion, its description, the rationale for its application, and the number of stations retained after each sequential screening step.

((Minor comment #7)) Line 60: Please rephrase--given the context is that the hydrology has undergone much human alteration, it's incongruous to call it a "natural laboratory" (but I agree with the sentiment!)

((Response #7)) We have replaced "a valuable natural laboratory" with "a valuable setting" to avoid implying natural conditions in river systems extensively modified by human activities.

((Minor comment #8)) Figure 1 caption appears to contain the Figure 4 caption too due to an error.

((Response #8)) Thank you for pointing out this error. We have corrected the Figure 1 caption.

((Minor comment #9)) Figure 1 domain appears to include parts of North Korea, so please alter wording accordingly (or change the figure). Namely, the wording is currently "Major river basins and climate zones across South Korea"--perhaps repeat the phrase "study domain"?

((Response #9)) We have replaced "across South Korea" with "across the study domain" in the Figure 1 caption to accurately reflect the spatial extent shown, which includes parts of North Korea.

((Minor comment #10)) Figure 3a. I think this panel is meant to show all the catchment

boundaries and how they fit into each other, but the low quality interferes. Please alter the figure to ensure all boundaries are visible.

((Response #10)) We have revised Figure 3a to improve its quality and ensure that all catchment boundaries are clearly visible.

((Minor comment #11)) Figure 3a. Please review the document to ensure the country naming is consistent. Since "South Korea" is used in the title, it should be used throughout, unlike Figure 3a in which the boundary is labelled "Korea".

((Response #11)) Thank you for this comment. We have revised the label in Figure 3a and reviewed the manuscript to ensure that "South Korea" is used consistently throughout.

((Minor comment #12)) Table 1. For the GLEAM PET, please specify the PET formulation relevant to the GLEAM data.

((Response #12)) We have revised Table 2 to specify the PET formulation relevant to the GLEAM data.

((Minor comment #13)) Figure 4. Regarding titles naming the variable being mapped: given there are 11 sub-plots, it is inconvenient for the reader to be forced to refer to the caption to learn what is being mapped in each case. Please add titles to each sub plot indicating the variable.

((Response #13)) We have added a variable label below each subplot in Figure 4 to allow readers to identify the mapped variables without referring to the caption. We placed these labels below the panels to maintain a balanced layout.

((Minor comment #14)) Table 2. Specify what is the difference between designed storage capacity and effective storage capacity.

((Response #14)) We have revised the descriptions of `dam_storage_capacity` and `dam_effective_storage` in Table 3 to clarify their distinction. The former represents the total design storage capacity, including storage unavailable for operational use, whereas the latter represents the usable storage capacity within the designated operating water-level range for water supply and operational management.

((Minor comment #15)) Line 343. Please change the order of material so that the motivation is stated prior to the specifics of the method. This is because the motivation often plays a role in what method is chosen. Please review the remainder of the manuscript for similar occurrences.

((Response #15)) We have reorganised the paragraph to state the motivation for including simulated streamflow products before introducing the LSTM and HBV models and explaining the model selection. The revised text first clarifies that these products complement the observed streamflow records.

((Minor comment #16)) Line 356. Given how short some of the discharge observations records are, surely an additional motivation is to be able to infill the short records backwards in time?

((Response #16)) We agree that extending short discharge records backwards in time is an additional motivation for providing simulated streamflow. We have revised the third purpose to explicitly include both filling gaps within observed records and extending the time series to periods before discharge observations began, where meteorological forcing data are available.

((Minor comment #17)) Figure 5. The performance of the models is relatively low. There's nothing wrong with that, but can the authors provide some initial comments as to why? Future users of this dataset will get similarly low results, but the difference is that they don't have the detailed local knowledge of the dataset authors to inform insights as to why. I note that the authors have provided this guidance already but only for the coastal catchments, which are a

small minority. What about the non-coastal catchments? Line 357 states potential reasons for low performance, but this is a general list provided prior to the modelling results--which of these turn out to be responsible? This discussion may be related to Fig. 7, which shows systematic issues with data (e.g. water balance issues) for some catchments. It is ok for this to be brief, as modelling results are not the subject of this paper.

((Response #17)) To respond to this comment, we have expanded the discussion accompanying Fig. 5 to address possible influences on model performance in non-coastal catchments, including human water management and uncertainties in the source data. We also refer to the water-balance departures discussed subsequently in Figure 7b, which highlight potential inconsistencies between catchment precipitation and discharge that may complicate rainfall–runoff modeling.