

<RESPONSE TO REFEREE #1>

- **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 sincerely thank Reviewer #1, Dr. Nikunj K. Mangukiya, for his insightful and constructive feedback. His comments have helped us improve the quality, clarity, and scientific rigor of our manuscript. Below, we address each comment in detail. All line numbers cited in our responses refer to the revised manuscript.

Specific Comment

((Comment #1)) The manuscript describes the 20% area discrepancy filtering criterion in Section 2.2, but nowhere does it specify the underlying digital elevation model processing, flow direction algorithm, or gauge snapping procedures used to delineate the 282 catchments from the 30-m ALOS DEM. The authors should explicitly document the delineation pipeline, software/tools used, snapping search radius tolerances, and how boundaries in flat coastal or highly urbanized alluvial plains were quality-checked against official national drainage boundaries.

((Response #1)) We agree that the original manuscript did not provide sufficient information regarding the catchment delineation procedure. We have therefore expanded Section 2.2 (Lines 127–137) 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 resulting catchments into vector boundaries.

((Comment #2)) In Figure 7b, multiple catchments exhibit long-term runoff coefficients substantially exceeding unity ($Q/P > 1.0$) or plot well below the lower Budyko demand limit ($E_{actual} \gg PET$). The statement on lines 441-443 ("Most catchments fall within the theoretical

water and energy limits...") overlooks these clear physical discrepancies. The authors must quantify the number of anomalous basins and investigate the underlying causes, such as orographic precipitation underestimation from ordinary kriging, trans-basin water transfers, deep inter-catchment groundwater flow, artificial drainage, or unmodeled dam/reservoir release regimes, and provide a quality/reliability flag in the dataset for basins with severe water balance closure errors.

((Response #2)) We recognize that the previous description understated the apparent water-balance inconsistencies in Figure 7b. We therefore re-examined the anomalous catchments and revised the corresponding discussion (Lines 447–454). 24 catchments fall below the theoretical energy-limit line, indicating potential water losses that cannot be explained by the standard Budyko water-balance assumptions alone. In addition, 16 catchments in urban areas exhibit runoff ratios greater than unity ($Q/P > 1$), indicating that observed streamflow exceeds precipitation over the analysis period. Such departures may be influenced by anthropogenic modifications, particularly interbasin water transfers, imported municipal water, and reservoir operations (see Section 4.4). These deviations may represent either genuinely non-closed water balances associated with human water management or inconsistencies among the source datasets.

((Comment #3)) The meteorological forcing dataset was developed using 2D Ordinary Kriging of ASOS/AWS weather stations onto a 0.1° grid. In topographically rugged terrain like the Korean Peninsula, standard spatial interpolation without elevation covariates (e.g., Kriging with External Drift, regression kriging, or PRISM-type elevation lapse rates) typically underestimates ridge-line and high-elevation precipitation. The authors should discuss whether elevation lapse rates were considered and compare the gridded precipitation against satellite/gauge-corrected benchmarks or local high-elevation gauges to assess potential precipitation underestimation.

((Response #3)) We acknowledge that elevation-based precipitation adjustments, such as elevation lapse rates, and kriging with external drift, were not explicitly incorporated into the meteorological forcing dataset. Instead, ordinary kriging was applied directly to the quality-controlled ASOS/AWS station observations.

The underlying gap-filling and spatial-interpolation framework was previously evaluated by Noh and Ahn (2022) through station-based validation and comparisons with several independent gridded precipitation products. Although these evaluations showed that the reconstructed dataset generally captured the broad spatial and temporal variability of precipitation over the Korean Peninsula, they did not specifically assess elevation-dependent biases using high-elevation gauges. Therefore, the existing validation cannot rule out localized precipitation underestimation in mountainous terrain.

However, we agree that ordinary kriging may smooth localized orographic precipitation gradients because it does not explicitly account for systematic relationships between precipitation and elevation. Consequently, precipitation may be underestimated along ridge lines and in other high-elevation areas. These uncertainties may propagate to precipitation-dependent hydrological signatures, water-balance analyses, and rainfall–runoff simulations. We have added a discussion of this limitation in Section 6.

((Comment #4)) A network of 282 catchments across South Korea contains significant spatial overlap and nested river networks. It would be beneficial for the users if authors can provide a topological routing matrix or metadata attributes (e.g., `is_headwater`, `upstream_basin_ids`, `nested_area_fraction`).

((Response #4)) We agree that explicit information on the topological relationships among the 282 catchments is important because many of the catchments overlap and are nested within the same river networks. Accordingly, we have added two new 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, whereas `upstream_basin_ids` provides the identifiers of the CAMELS-KR catchments located upstream of each catchment. The corresponding descriptions of these attributes have been added to Section 4.1, Table 3, and the dataset documentation.

((Comment #5)) Observed streamflow records vary substantially in length (10 to 35 years; Fig 3c) and include non-continuous gaps. The authors should clarify whether hydrologic signatures

were computed over the full variable record or harmonized common benchmark window, and address how structural regime shifts (e.g., post-2010 construction from the Four Major River Project) affect static signature stability.

((Response #5)) We thank the reviewer for raising this important point. The hydrologic signatures were calculated using all valid observations within the full available streamflow record for each catchment. In addition, we agree that structural changes in river regulation may affect the temporal stability of hydrologic signatures. In particular, the Four Major Rivers Project, implemented primarily between 2009 and 2012, included the construction and operation of 16 multifunctional weirs, channel dredging, and modifications to water-storage and regulation infrastructure. These interventions may have altered water levels, low-flow conditions, flow seasonality, and stage–discharge relationships at stations located along the regulated main stems or downstream of the affected infrastructure. However, these effects are spatially heterogeneous and should not be assumed to apply equally to all 282 catchments.

We note that users interested in natural flow behavior or temporal change should consider the catchment-specific observation periods, data gaps, and degree of river regulation when selecting catchments and interpreting the signatures. A systematic change-point analysis and attribution of individual regime shifts to specific infrastructure are beyond the scope of the present dataset study.

((Comment #6)) While soil characteristics from SoilGrid 2.0 are well documented, the dataset lacks attributes describing underlying bedrock geology, lithology, and hydrogeological properties (e.g., hydraulic conductivity and porosity from GLHYMPS or global lithological datasets like GLiM). Incorporating standard lithological classes and bedrock permeability attributes would bring CAMELS-KR into complete alignment with other CAMELS datasets.

((Response #6)) We thank the reviewer for this suggestion. We agree that geological and hydrogeological properties are important catchment descriptors because they influence subsurface water storage, groundwater movement, and runoff generation. In response, we derived and added nine new catchment-scale attributes from GLHYMPS and GLiM.

Two hydrogeological attributes were obtained from GLHYMPS: `geo_porosity`, representing

catchment-averaged subsurface porosity, and `geo_log10_permeability`, representing the catchment-averaged base-10 logarithm of subsurface permeability. We also calculated the catchment-area percentages of the seven GLiM lithological classes occurring within the CAMELS-KR catchments: acid plutonic rocks (`pa_perc`), metamorphic rocks (`mt_perc`), acid volcanic rocks (`va_perc`), siliciclastic sedimentary rocks (`ss_perc`), basic plutonic rocks (`pb_perc`), basic volcanic rocks (`vb_perc`), and carbonate sedimentary rocks (`sc_perc`). These nine attributes are now included in the revised CAMELS-KR dataset and described in the revised manuscript and dataset documentation.

((Comment #7)) The streamflow reconstruction section requires additional technical specifics. For the regional LSTM model, which specific static attributes and meteorological forcings were fed into the feature vector? was hyperparameter tuning performed via k-fold cross-validation or an out-of-basin spatial split? For the HBV model, which optimization algorithm and objective function formulation were applied during local calibration? what parameter search ranges were defined? The authors can also include diagnostic attribution to clearly show where and why model performance degrades.

((Response #7)) To respond to this comment, we have expanded Section 5.2 to describe the input variables, hyperparameter-selection procedure, HBV calibration algorithm, objective function, and parameter search ranges. Specifically, we have clarified that the regional LSTM used eight daily meteorological variables—precipitation, minimum and maximum temperature, relative humidity, mean and maximum wind speed, solar radiation, and Penman–Monteith-based PET—together with static catchment attributes representing topography, land cover, and soil characteristics. The LSTM hyperparameters were predetermined through a preliminary grid search before the final regional training. the HBV model was calibrated independently for each catchment using the genetic algorithm implemented in the `pymoo` package. The calibration objective was the MSE between observed and simulated daily streamflow. The HBV parameter search ranges followed Feng et al. (2022).

((Comment #8)) Section 8 (Data Availability) states that daily forcing includes vapor pressure and sunshine duration, but these variables are not present in the dataset.

((**Response #8**)) We thank the reviewer for identifying this inconsistency. We have removed these two variables from the list of daily meteorological forcing variables in Section 8.

((**Comment #9**)) The caption for Figure 1 (lines 106–111) erroneously includes a duplicated description belonging to Figure 4.

((**Response #9**)) Thank you for pointing out our mistake. The caption for Figure 1 has been corrected in the revised manuscript.

((**Comment #10**)) The time series data include negative water level values for certain gauges, but the manuscript lacks an explanation of the underlying datum definition beyond describing it in Table 1 as "relative to the station-specific zero datum". In hydrometric networks, negative readings typically arise from channel bed degradation/scour below an established gauge zero or arbitrary local datum definitions. The authors should explicitly clarify in Section 3.1 why negative stages occur, state whether gauge zeros have shifted over time, and include the official gauge zero datum elevation (in above mean sea level or the national geodetic datum) in the static attribute table so users can convert relative stage into absolute hydraulic head.

((**Response #10**)) The point is well taken. The original `water_level` variable represents the daily stage relative to the station-specific gauge zero. Negative values occur when the observed water surface is below this local reference level and therefore do not necessarily indicate erroneous observations.

To facilitate the use of the water-level records in hydraulic and cross-catchment analyses, we have added a new daily variable, `water_surface_level`, representing the water-surface elevation relative to mean sea level (m a.s.l.). This variable was calculated as follows:

$$\text{water_surface_level}_{i,t} = \text{water_level}_{i,t} + \text{gauge_zero_elevation}_i \quad \text{Eq. R1}$$

Here, `gauge_zero_elevationi` denotes the official elevation of the gauge zero relative to mean

sea level at i th station. We have revised Section 3.1 and Table 2 to clarify the occurrence of negative relative stages and the definition and calculation of `water_surface_level`.

((Comment #11)) In Figure 8a (number of dams and reservoirs), the color ramp spans up to ~ 800 , but the inset histogram shows that the overwhelming majority of catchments have counts concentrated below 0–100. As a result, almost all catchment points appear uniformly pale yellow with little to no visible distinction across moderate differences. Similar issue is there in Figure 8b. Also, in Figure 8c, the colorbar map represent population density. The unit on the colorbar is labeled as $[\text{km}^{-2}]$, which is missing the actual quantity (e.g., $[\text{persons km}^{-2}]$).

((Response #11)) We acknowledge that the original linear color scales obscured differences among catchments with low to moderate numbers of dams and reservoirs. We have therefore revised Figures 8a and 8b to use logarithmic color scales, thereby improving the visual distinction across the predominant range of values. In addition, we have corrected the unit of population density in Figure 8c from $[\text{km}^{-2}]$ to $[\text{persons km}^{-2}]$.

((Reference))

Feng, D., Liu, J., Lawson, K., & Shen, C. (2022). Differentiable, learnable, regionalized process-based models with multiphysical outputs can approach state-of-the-art hydrologic prediction accuracy. *Water Resources Research*, 58(10), e2022WR032404.