

Caravan MultiMet: Extending Caravan with Multiple Weather Nowcasts and Forecasts

Guy Shalev¹ and Frederik Kratzert¹

¹Google Research

Correspondence: Guy Shalev (guysha@google.com)

Abstract. The Caravan large-sample hydrology dataset (Kratzert et al., 2023) was created to standardize and harmonize stream-flow data from various regional datasets, combined with globally available meteorological forcing and catchment attributes. This community-driven project also allows researchers to conveniently extend the dataset for additional basins, as done 6 times to date (see <https://github.com/kratzert/Caravan/discussions/10>). We present a novel extension to Caravan, focusing on enriching the meteorological forcing data. Our extension adds three precipitation nowcast products (CPC, IMERG v07 Early, and CHIRPS) and three weather forecast products (ECMWF IFS HRES, GraphCast, and CHIRPS-GEFS) to the existing ERA5-Land reanalysis data. The inclusion of diverse data sources, particularly weather forecasts, enables more robust evaluation and benchmarking of hydrological models, especially for real-time forecasting scenarios. To the best of our knowledge, this extension makes Caravan the first large-sample hydrology dataset to incorporate weather forecast data, significantly enhancing its capabilities and fostering advancements in hydrological research, benchmarking, and real-time hydrologic forecasting. The data is publicly available under a CC-BY-4.0 license on Zenodo in two parts (<https://zenodo.org/records/14161235>, <https://zenodo.org/records/14161281>) and on Google Cloud Platform (GCP) – see more under the Data Availability chapter.

1 Introduction

Large-sample hydrology (LSH) is crucial for understanding global water resources and addressing challenges like climate change impacts, flood and drought forecasting, and water management (Addor et al., 2020). Several LSH datasets have been instrumental in advancing hydrologic research. Notably, MOPEX (Andréassian et al., 2006) offered valuable hydrometeorological data for diverse catchments, contributing to model development and evaluation. Newman et al. (2015) in combination with Addor et al. (2017) with the release of the CAMELS dataset for the contiguous United States sparked a series of publications for regional LSH datasets. As of today, the existing dataset are CAMELS-AUS (Fowler et al., 2021), CAMELS-BR (Chagas et al., 2020), CAMELS-CH (Höge et al., 2023), CAMELS-CL (Alvarez-Garreton et al., 2018), CAMELS-DE (Loritz et al., 2024), CAMELS-DK (Liu et al., 2024), CAMELS-FR (Delaique et al., 2024), CAMELS-GB (Coxon et al., 2020), and CAMELS-INDIA (Mangukiya et al., 2024). Worth mentioning are also the LamaH-CE (Klingler et al., 2021) and LamaH-ICE (Helgason and Nijssen, 2024) datasets, which in contrast to the CAMELS datasets additionally provide information about the gauge connectivity. However, combining and comparing these datasets is not straightforward (Addor et al., 2020), as usually these regional datasets make use of high resolution, only locally available, forcing data and attribute maps. Here,

the recent introduction of Caravan has marked a significant step forward – by uniting and standardizing data from several large-sample hydrology datasets, and allowing community contributions to expand it to even more regions (see updating list here <https://github.com/kratzert/Caravan/discussions/10> including Koch (2022), Morin (2023), Höge et al. (2023), Casado Rodríguez (2023), Helgason (2024), and Färber et al. (2023)).

30 The forcing timeseries available in Caravan are derived from ERA5-Land (Muñoz-Sabater et al., 2021). This was the product of choice due to its global coverage, relatively high quality, various surface variables, long records, availability on Earth Engine, and wide acceptance within the hydrologic community. While ERA5-Land has been invaluable for hydrologic modeling, it presents certain limitations – as with any global product, the accuracy and reliability can vary spatially and temporally depending on the density and quality of the underlying observations. It has been shown that using a combination of several
35 weather sources can increase performance of data driven hydrologic models (Kratzert et al. (2021)). This increases the motivation not to solely rely on a single product when possible, as also done operationally by Google’s flood forecasting model (See <https://g.co/floodhub>).

Furthermore, relying solely on reanalysis data restricts hydrologic modeling to hindcasting scenarios. To advance hydrologic forecasting and to enable model intercomparison studies in the context of operational forecasting applications, it is essential to
40 incorporate a historic archive of real-time weather forecast data. This allows researchers to evaluate model performance under realistic conditions, assess forecast uncertainty propagation through hydrologic models, and ultimately improve prediction accuracy for water resource management and hazard mitigation.

By including diverse weather data products and weather forecasts, this extension to Caravan bridges a critical gap in large-sample hydrology. It provides a unique platform for benchmarking hydrologic models in operational settings (Nearing
45 et al. (2024) and more recently <https://research.google/blog/a-flood-forecasting-ai-model-trained-and-evaluated-globally/>), enabling researchers to rigorously evaluate model performance across a range of meteorological forcings and forecast horizons. This capability is paramount for developing and deploying reliable hydrologic forecasting systems that can inform decision-making in real-time, ultimately enhancing our resilience to water-related challenges.

2 Data

50 Caravan includes meteorological forcing data from ERA5-Land, aggregated to daily resolution in the local time of each gauge location. You can find all details of the processing of the meteorological data in Kratzert et al. (2023). In agreement with the existing data in Caravan, we spatially average all weather data across the catchment area and aggregate to daily resolution. However, for this extension, we keep all data in its original time (which is UTC+0) for multiple reasons: First, not all data products are available at sub-daily resolution, which is required to shift the data to local time, before computing daily statistics.
55 Second, UTC+0 is the standard for global weather products, and by keeping the data of this extension in UTC+0, we make it easier to compare to additional products that are not included in this extension. For these reasons, our extension also includes daily aggregates of ERA5-Land, this time computed in UTC+0 and not in local time, as in the original Caravan dataset.

Since not all of this data is available on Earth Engine, we process this data using internal Google infrastructure for all existing Caravan gauges, including all extensions. This approach has the limitation of not covering future community contributions. We
60 aspire to mitigate this downside by periodical updates to the dataset.

2.1 Structure of the dataset

For each nowcast or reanalysis product, we provide a single zarr file indexed by (1) basin and (2) date. And, for each forecast product, we provide a single zarr file which is indexed by (1) basin, (2) date, and (3) lead-time. The zarr format allows to efficiently read only parts of the dataset (e.g., only a subset of basins, dates, and lead-times) without the need to read all data
65 into memory.

The data for the nowcast product is left-labeled, meaning that the value that appears under “date D”, refers to the time period [D 00:00:00, D+1 00:00:00]. For the forecast dataset, the date is the instant where the forecast was supposedly issued – the fact these are sometimes reforecasts, and that operational forecasts have a few hours delay is ignored. The lead-times are expressed by `datetime.timedelta(days=X)`. The first value is `datetime.timedelta(days=1)`, and refers to the data in the range [D 00:00:00, D+1 00:00:00]; the second lead-time is indexed as `datetime.timedelta(days=2)`, and refers to [D+1 00:00:00, D+2 00:00:00];
70 and so on. For instance, if one would want to compare the data from all products for some date, they could query for date D, and additionally query for the first lead-time if it is a forecast product.

2.2 Missing Data

The availability of each data-source was limited by the original data-sources, or the version of the data that Google possessed
75 at the time. This can include missing dates (as in CHIRPS-GEFS), partial spatial cover (as in CHIRPS and CHIRPS-GEFS), and partial variables and years (as in HRES). We highlight the specific limitations in the relevant chapters below. Missing data-points are always denoted with nan values. In case a basin had no valid values for some weather product, it is removed (but will still exist for other products).

2.3 Spatial Coverage

80 This extension covers 22,492 gauges included in the original Caravan dataset and its extensions, across 48 countries. See figure 1.

3 Nowcast Products

3.1 CPC

The Climate Prediction Center (CPC) Unified Gauge-Based Analysis of Daily Precipitation is a gridded precipitation product
85 generated by NOAA’s Climate Prediction Center. This dataset is part of the CPC Unified Precipitation Project, which aims to produce a suite of high-quality, consistent precipitation products by integrating various data sources using optimal interpolation

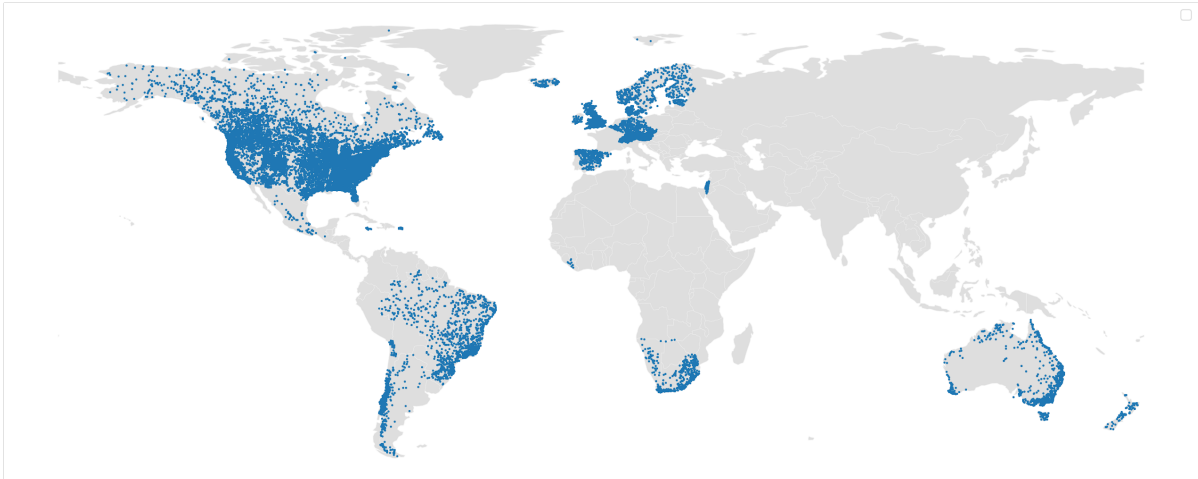


Figure 1. The locations of Caravan gauges covered by the Caravan MultiMet extension.

(OI) techniques (Xie et al., 2007). The original temporal resolution is daily, and the spatial resolution is 0.5°. You can read more about the underlying dataset at <https://psl.noaa.gov/data/gridded/data.cpc.globalprecip.html> .

The data provided in the Caravan-MultiMet extension covers the range of [1979-01-01, 2024-07-31]. Sporadic dates and 90 locations where the data is missing in the original dataset are filled with nan values (see figure 2).

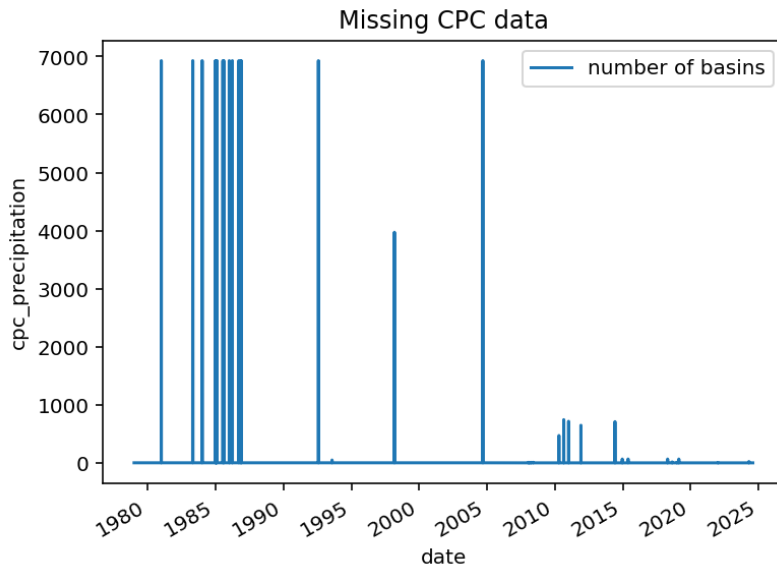


Figure 2. The number of basins with missing CPC data by date.

Product	Variable	Units	Availability	Lead-times
CPC	precipitation	<i>mm</i>	[1979-01-01, 2024-07-31] with sporadic missing data	-
IMERG v07 Early	precipitation	<i>mm</i>	[2000-05-01, 2024-10-31]	-
CHIRPS	precipitation	<i>mm</i>	[1981-01-01, 2024-07-30] within latitudes [-50°, 50°]	-
ERA5-Land	dewpoint_temperature_2m potential_evaporation_DEPRECATED potential_evaporation_FAO_PENMAN_MONTEITH snow_depth_water_equivalent surface_net_solar_radiation surface_net_thermal_radiation surface_pressure temperature_2m total_precipitation u_component_of_wind_10m v_component_of_wind_10m volumetric_soil_water_layer_1 volumetric_soil_water_layer_2 volumetric_soil_water_layer_3 volumetric_soil_water_layer_4	°C <i>mm</i> <i>mm</i> <i>mm</i> Wm^{-2} Wm^{-2} <i>kPa</i> °C <i>mm</i> ms^{-1} ms^{-1} m^3/m^3 m^3/m^3 m^3/m^3 m^3/m^3	[1950-01-01, 2024-10-31]	-
CHIRPS-GEFS	precipitation	<i>mm</i>	[2000-01-01, 2024-01-31] with some missing dates. within latitudes [-50°, 50°]	1-16 days
ECMWF HRES	surface_net_solar_radiation surface_net_thermal_radiation surface_pressure temperature_2m total_precipitation	Wm^{-2} Wm^{-2} <i>kPa</i> °C <i>mm</i>	[2012-07-01, 2024-06-30]	1-10 days
GraphCast	temperature_2m total_precipitation u_component_of_wind_10m v_component_of_wind_10m	°C <i>mm</i> ms^{-1} ms^{-1}	[2016-01-02, 2023-12-21]	1-10 days

Table 1. Summary of all available data in Caravan MultiMet

3.2 IMERG

The Integrated Multi-satellitE Retrievals for GPM (IMERG) is a NASA product estimating global surface precipitation rates at a spatial resolution of 0.1° every half-hour, starting in 2000. IMERG has three Runs with varying latency in response to a range of application needs: rapid-response applications (Early Run, 4-h latency), same/next-day applications (Late Run, 14-h
95 latency), and post-real-time research (Final Run, 3.5-month latency). In this dataset we use the “Early” version of IMERG which is most applicable for time-sensitive real-time forecasting. Version 07 of IMERG was released in June 2024, and is to date the current version of the dataset. More information is available at:

https://data.nasa.gov/dataset/GPM-IMERG-Early-Precipitation-L3-Half-Hourly-0-1-d/mek8-8ssg/about_data. The data was obtained from <https://jsimpsonhttps.pps.eosdis.nasa.gov/imerg/early/>.

100 The data provided in the Caravan-MultiMet extension covers the range of [2000-05-01, 2024-10-31], with no missing data steps.

3.3 CHIRPS

CHIRPS (Climate Hazards group Infrared Precipitation with Stations) is a quasi-global high resolution, daily, precipitation dataset. Developed by the Climate Hazards Center at the University of California, Santa Barbara, CHIRPS blends satellite
105 imagery with in-situ station data to create gridded rainfall estimates. More information is available at: <https://www.chc.ucsb.edu/data/chirps>. The spatial resolution of the data is 0.05° , making this the highest spatial resolution data in this extension. Unfortunately, CHIRPS only covers the latitude range [-50, 50]. We provide the processed data for Caravan basins where the entire catchment is within this covered region.

110 The data provided in the Caravan-MultiMet extension covers the range of [1981-01-01, 2024-07-30], with no missing data steps in this range.

3.4 ERA5-Land recomputed at UTC daily aggregation

ERA5-Land provides hourly high resolution information for several surface variables. You can read more about ERA5-Land and the way it is processed in the original Caravan paper.

115 We recompute the ERA5-Land data provided in the original Caravan dataset with one major difference – the data is aggregated into daily resolution in UTC timezone for all basins, as opposed to the original dataset which computed the daily aggregation after shifting to the local timezone of each basin. This makes the ERA5-Land data in this extension compatible with the other meteorological datasets in Caravan-MultiMet, where it is not always possible to shift to the local timezone (e.g. in the case of native daily resolution products). It is important to note that the discharge data in Caravan is typically affected by the local timezone and does not necessarily match UTC. Another difference is that we only include daily averages (or daily
120 sums, in the case of accumulating features of ERA5-Land), as opposed to the original Caravan dataset, which also provided the daily min and max values.

Several studies raised concerns regarding the reliability of the ERA5-Land potential evaporation (Clerc-Schwarzenbach et al. (2024), Kratzert et al. (2023), Klingler et al. (2021)). We therefore add an additional potential evaporation variable, based on the FAO Penman-Monteith (Allen et al., 1998) using other ERA5-Land variables, at a daily resolution, based on the code provided by Singer et al. We name the new variable "potential_evaporation_FAO_PENMAN_MONTEITH", and include the original ERA5-Land variable (renamed to "potential_evaporation_DEPRECATED") to allow for their comparison.

The data provided in the Caravan-MultiMet extension covers the range of [1950-01-01, 2024-10-31], with no missing data steps for the provided basins.

4 Forecast Products

130 4.1 CHIRPS-GEFS

CHIRPS-GEFS is a bias-corrected and downscaled version of NCEP Global Ensemble Forecast System (GEFS) version 12 precipitation forecasts made to be spatially compatible with Climate Hazards Center InfraRed Precipitation with Stations (CHIRPS) data. It provides forecasts issued daily, with lead-times from 1 day to 16 days in daily resolution. The spatial resolution is 0.05° (as in CHIRPS). Unfortunately, as in CHIRPS, this product only covers the latitude range $[-50, 50]$. We provide the processed data for Caravan basins where the entire catchment is within this covered region. The data provided in the Caravan-MultiMet extension covers the range of [2000-01-01, 2024-01-31], with some missing data steps (see figure 3).

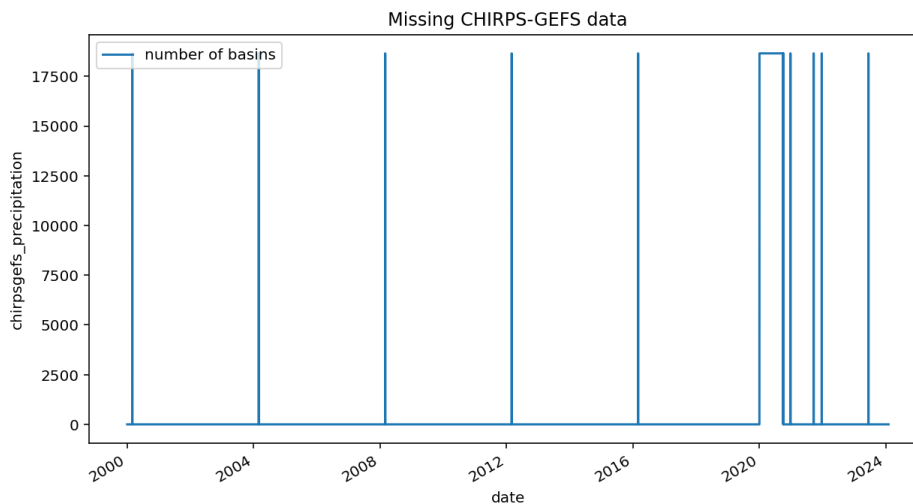


Figure 3. The number of basins with missing CHIRPS-GEFS data by date. Notably, a large part of 2020 is missing for all basins, and February 29th on leap years.

4.2 ECMWF IFS HRES

The European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecasting System (IFS) High-Resolution (HRES) model provides deterministic weather forecasts which include several variables. ECMWF's global forecasts are recognised as amongst the most skilful in the world. It operates at an approximate spatial resolution of 0.1° and provides a weather forecast for the next 10 days.

Acquiring ECMWF's forecast data is not as straightforward as other products. On the one hand, this makes this part of the extension more valuable, saving the user a lot of work acquiring the data. On the other hand, the limitations we had ourselves in querying the data affected the data we can responsibly provide in this dataset – thus limiting us to only 5 variables, and only several years of data even though the full archive includes more.

The underlying data used to create this dataset is based on forecasts with issue time 00:00 UTC. The lead-times available are hourly for the first 90 hours; 3-hourly for hours 90-144; and 6 hourly onwards until 240 hours after the issue time. All of these are aggregated to daily values.

The surface variables provided in this extension are a subset of the ERA5-Land set of variables, and are in many ways compatible and comparable with that data – see the list of variables and their units in Table 1. The data provided in the Caravan-MultiMet extension covers the range of [2016-01-01, 2024-09-30], with no missing data steps.

4.3 GraphCast

GraphCast is a groundbreaking machine learning (ML) model for global weather forecasting developed by DeepMind (Lam et al., 2023). The model leverages graph neural networks (GNNs) to process spatially structured weather data and generate forecasts at a spatial resolution of 0.1° degrees, with 10 days lead time, at a 6-hourly temporal resolution.

Although the data is available for 4 issue-times per day, we use only the one produced at 00:00. The lead-times of GraphCast have a 6-hour density – we aggregate the 4 lead-times of each day (by mean or sum, depending on the variable) to produce the daily lead-time density for 10 days. The surface variables provided in this extension are precipitation, temperature, and wind (u10 and v10 components). Similarly to HRES, these too are compatible with the ERA5-Land data.

This version of GraphCast has been generated by the DeepMind group by finetuning and initializing the model with HRES as the initial state, as opposed to ERA5 in the public version. This means that this data is representative the quality the GraphCast model can generate in real-time. The data for each year was generated using a GraphCast model trained until that year, which could potentially cause very minor inconsistencies and discontinuities between the forecasts of different years.

The data provided in the Caravan-MultiMet extension covers the range of [2016-01-02, 2023-12-21], with no missing data steps.

5 Data Verification

5.1 ERA5-Land agreement with the original Caravan

As a first step, we checked that our reprocessing of ERA5-Land agrees with the original dataset. Small and insignificant discrepancies are probably produced due to the processing code – which is not entirely identical as the one used on Earth Engine. Since the original Caravan dataset was shifted to local timezone, we focus our analysis on an arbitrary gauge in Great Britain with UTC timezone, "camelsgb_54007" in the year 2000. Note that the original Caravan also rounds the values to 2 decimal digits, while this extension doesn't. We apply the comparison a rounded version of the extension data. The results are shared in Table 2. Figure 4 shows the data for total_precipitation, the variable with the strongest disagreement ($R^2 = 0.988$) to get a feeling of the magnitude of such differences.

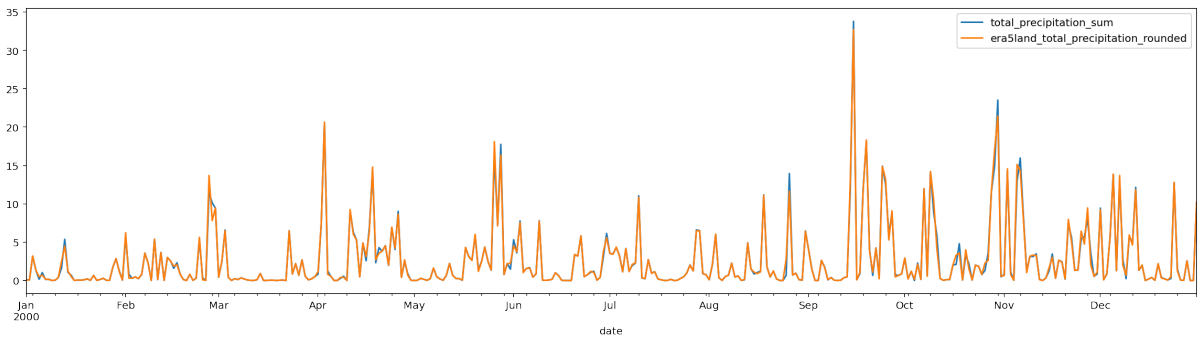


Figure 4. The ERA5-Land total precipitation for a year at the gauge "camelsgb_54007", which shows strong agreement between the two versions of the data – blue for the original Caravan dataset, orange for the MultiMet extension. It seems they have occasional disagreements on peak values.

5.2 Verification of other products

Now that we have verified the ERA5-Land data of the extension agrees with the original dataset, we use it for reference and compare it to the other products. For each other variable in the other products, there is a trivially matching ERA5-Land variable (most commonly this is total_precipitation, but for HRES and GraphCast several others exist). We compare each variable to their matching ERA5-Land variable, by applying R^2 score with ERA5-Land taken as the ground truth – results are presented in Table 3. We also apply this test with shifts to the ERA5 data, to make sure we do not have an off-by-one error in the process. Here, the maximal R^2 is obtained with no shifts for all variables. The reasonably high R^2 scores for the correct shift also provides verification that the units are correct, a property that is not captured by metrics such as correlation. The score was computed on all dates that where both variables were available, and we include the number of samples used for the computation. The results are presented in Table 3.

Variable	R^2
dewpoint_temperature_2m_mean	0.999
temperature_2m_mean	0.999
potential_evaporation_sum	0.999
snow_depth_water_equivalent_mean	0.999
surface_pressure_mean	0.998
surface_net_solar_radiation_mean	1.0
surface_net_thermal_radiation_mean	0.994
volumetric_soil_water_layer_1_mean	0.996
volumetric_soil_water_layer_2_mean	0.998
volumetric_soil_water_layer_3_mean	0.999
volumetric_soil_water_layer_4_mean	0.988
total_precipitation_sum	0.989
u_component_of_wind_10m_mean	0.998
v_component_of_wind_10m_mean	0.998

Table 2. R^2 score of all ERA5-Land variables for the year 2000 compared to the data in the original Caravan dataset. This is a verification that the data is in the correct units, and that the differences in the processing are of a reasonable size. This analysis is over the arbitrary gauge with UTC timezone "camelsgb_54007"

185 6 Data and Code Availability

The data is available for download on Zenodo in two parts (<https://zenodo.org/records/14161235>, <https://zenodo.org/records/14161281>).

190 Additionally, a colab with a usage example of the dataset is available at <https://github.com/kratzert/Caravan/>. The colab reads the data from zarr files stored on GCP, which allows accessing only subsets of the data (specific variables, basins, dates, and lead-times) without the need to download the entire dataset and to read all of it into memory.

The Caravan-MultiMet dataset is released under CC-BY-4.0 license, like the majority of Caravan. The original license and source of each individual product are attached in the dataset itself. The original Caravan dataset is available at <https://doi.org/10.5281/zenodo.7540792>.

Any feedback, especially concerning issues with the data itself, will be highly appreciated.

Variable	R^2 shift -1	R^2 no shift	R^2 shift +1	Num. valid samples for computation
cpc_precipitation	-0.29	0.49	-0.08	16623
imerg_precipitation	-0.84	0.1	-1.17	8919
chirps_precipitation	-0.39	0.14	-0.69	15917
chirpsgefs_precipitation	-0.74	0.5	-0.69	8512
hres_surface_net_solar_radiation	0.8	0.98	0.78	3196
hres_surface_net_thermal_radiation	0.66	0.97	0.64	3196
hres_surface_pressure	0.53	0.85	0.53	3196
hres_temperature_2m	0.96	1.0	0.96	3196
hres_total_precipitation	-0.31	0.84	-0.33	3196
graphcast_temperature_2m	0.95	0.98	0.94	2911
graphcast_total_precipitation	-0.13	0.87	-0.1	2911
graphcast_u_component_of_wind_10m	0.24	0.87	0.16	2911
graphcast_v_component_of_wind_10m	0.23	0.88	-0.02	2911

Table 3. R^2 score of all non-ERA5-Land variables with their matching ERA5-Land variable, with and without shifts. This is a verification that the data is not shifted, and that the different variables agree to a reasonable extent. This analysis was done for the sample gauge "GRDC_6217400"

195 References

- Addor, N., Newman, A. J., Mizukami, N., and Clark, M. P.: The CAMELS data set: catchment attributes and meteorology for large-sample studies, *Hydrology and Earth System Sciences*, 21, 5293–5313, 2017.
- Addor, N., Do, H. X., Alvarez-Garreton, C., Coxon, G., Fowler, K., and Mendoza, P. A.: Large-sample hydrology: recent progress, guidelines for new datasets and grand challenges, *Hydrological Sciences Journal*, 65, 712–725, <https://doi.org/10.1080/02626667.2019.1683182>, 2020.
- Allen, R. G., Pereira, L. S., Raes, D., and Smith, M.: Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56, Fao, Rome 300, D05109, <https://www.fao.org/4/x0490e/x0490e05.htm>, 1998.
- Alvarez-Garreton, C., Mendoza, P. A., Boisier, J. P., Addor, N., Galleguillos, M., Zambrano-Bigiarini, M., Lara, A., Puelma, C., Cortes, G., Garreaud, R., et al.: The CAMELS-CL dataset: catchment attributes and meteorology for large sample studies—Chile dataset, *Hydrology and Earth System Sciences*, 22, 5817–5846, 2018.
- Andréassian, V., Hall, A., Chahinian, N., and Schaake, J.: Large Sample Basin Experiments for Hydrological Model Parameterization: Results of the Model Parameter Experiment — MOPEX, *IAHS Publ.*, 307, 313–338, <https://doi.org/10.1080/13241583.2007.11465316>, 2006.
- Casado Rodríguez, J.: CAMELS-ES: Catchment Attributes and Meteorology for Large-Sample Studies – Spain, <https://doi.org/10.5281/zenodo.8373020>, 2023.
- Chagas, V. B., Chaffe, P. L., Addor, N., Fan, F. M., Fleischmann, A. S., Paiva, R. C., and Siqueira, V. A.: CAMELS-BR: hydrometeorological time series and landscape attributes for 897 catchments in Brazil, *Earth System Science Data*, 12, 2075–2096, 2020.

- Clerc-Schwarzenbach, F. M., Selleri, G., Neri, M., Toth, E., van Meerveld, I., and Seibert, J.: HESS Opinions: A few camels or a whole caravan?, *EGUosphere*, 2024, 1–29, <https://doi.org/10.5194/egusphere-2024-864>, 2024.
- 215 Coxon, G., Addor, N., Bloomfield, J. P., Freer, J., Fry, M., Hannaford, J., Howden, N. J., Lane, R., Lewis, M., Robinson, E. L., et al.: CAMELS-GB: Hydrometeorological time series and landscape attributes for 671 catchments in Great Britain, *Earth System Science Data*, 12, 2459–2483, 2020.
- Delaigue, O., Guimarães, G. M., Brigode, P., Génot, B., Perrin, C., Soubeyroux, J.-M., Janet, B., Addor, N., and Andréassian, V.: CAMELS-FR dataset: A large-sample hydroclimatic dataset for France to explore hydrological diversity and support model benchmarking, *Earth System Science Data Discussions*, 2024, 1–27, <https://doi.org/10.5194/essd-2024-415>, 2024.
- 220 Fowler, K. J., Acharya, S. C., Addor, N., Chou, C., and Peel, M. C.: CAMELS-AUS: hydrometeorological time series and landscape attributes for 222 catchments in Australia, *Earth System Science Data*, 13, 3847–3867, 2021.
- Färber, C., Plessow, H., Kratzert, F., Addor, N., Shalev, G., and Looser, U.: GRDC-Caravan: extending the original dataset with data from the Global Runoff Data Centre, <https://doi.org/10.5281/zenodo.10074416>, 2023.
- 225 Helgason, H. B., B. N.: LamaH-Ice: LARge-SaMple DATA for Hydrology and Environmental Sciences for Iceland, *HydroShare*, <https://doi.org/https://doi.org/10.4211/hs.86117a5f36cc4b7c90a5d54e18161c91>, 2024.
- Helgason, H. B. and Nijssen, B.: LamaH-Ice: LARge-SaMple DATA for Hydrology and Environmental Sciences for Iceland, *Earth System Science Data*, 16, 2741–2771, <https://doi.org/10.5194/essd-16-2741-2024>, 2024.
- Höge, M., Kauzlaric, M., Siber, R., Schönenberger, U., Horton, P., Schwanbeck, J., Floriancic, M. G., Viviroli, D., Wilhelm, S., Sikorska-Senoner, A. E., Addor, N., Brunner, M., Pool, S., Zappa, M., and Fenicia, F.: CAMELS-CH: hydro-meteorological time series and landscape attributes for 331 catchments in hydrologic Switzerland, *Earth System Science Data*, 15, 5755–5784, <https://doi.org/10.5194/essd-15-5755-2023>, 2023.
- 230 Höge, M., Kauzlaric, M., Siber, R., Schönenberger, U., Horton, P., Schwanbeck, J., Floriancic, M. G., Viviroli, D., Wilhelm, S., Sikorska-Senoner, A. E., Addor, N., Brunner, M., Pool, S., Zappa, M., and Fenicia, F.: Catchment attributes and hydro-meteorological time series for large-sample studies across hydrologic Switzerland (CAMELS-CH), <https://doi.org/10.5281/zenodo.7928595>, 2023.
- 235 Klingler, C., Schulz, K., and Herrnegger, M.: LamaH-CE: LARge-SaMple DATA for Hydrology and Environmental Sciences for Central Europe, *Earth System Science Data*, 13, 4529–4565, 2021.
- Koch, J.: Caravan extension Denmark - Danish dataset for large-sample hydrology, <https://doi.org/10.5281/zenodo.7396466>, 2022.
- Kratzert, F., Klotz, D., Hochreiter, S., and Nearing, G. S.: A note on leveraging synergy in multiple meteorological data sets with deep learning for rainfall–runoff modeling, *Hydrology and Earth System Sciences*, 25, 2685–2703, <https://doi.org/10.5194/hess-25-2685-2021>, 2021.
- 240 Kratzert, F., Nearing, G., Addor, N., Erickson, T., Gauch, M., Gilon, O., Gudmundsson, L., Hassidim, A., Klotz, D., Nevo, S., et al.: Caravan-A global community dataset for large-sample hydrology, *Scientific Data*, 10, 61, 2023.
- Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., Ewalds, T., Eaton-Rosen, Z., Hu, W., Merose, A., Hoyer, S., Holland, G., Vinyals, O., Stott, J., Pritzel, A., Mohamed, S., and Battaglia, P.: Learning skillful medium-range global weather forecasting, *Science*, 382, 1416–1421, <https://doi.org/10.1126/science.adi2336>, 2023.
- 245 Liu, J., Koch, J., Stisen, S., Troldborg, L., Højberg, A. L., Thodsen, H., Hansen, M. F. T., and Schneider, R. J. M.: CAMELS-DK: Hydrometeorological Time Series and Landscape Attributes for 3330 Catchments in Denmark, *Earth System Science Data Discussions*, 2024, 1–30, <https://doi.org/10.5194/essd-2024-292>, 2024.

- 250 Loritz, R., Dolich, A., Acuña Espinoza, E., Ebeling, P., Guse, B., Götte, J., Hassler, S. K., Hauffe, C., Heidbüchel, I., Kiesel, J., Mälicke, M., Müller-Thomy, H., Stölzle, M., and Tarasova, L.: CAMELS-DE: hydro-meteorological time series and attributes for 1555 catchments in Germany, *Earth System Science Data Discussions*, 2024, 1–30, <https://doi.org/10.5194/essd-2024-318>, 2024.
- Mangukiya, N. K., Kumar, K. B., Dey, P., Sharma, S., Bejagam, V., Mujumdar, P. P., and Sharma, A.: CAMELS-INDIA: hydrometeorological time series and catchment attributes for 472 catchments in Peninsular India, *Earth System Science Data Discussions*, 2024, 1–43, <https://doi.org/10.5194/essd-2024-379>, 2024.
- 255 Morin, E.: Caravan extension Israel - Israel dataset for large-sample hydrology, <https://doi.org/10.5281/zenodo.7758516>, 2023.
- Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., et al.: ERA5-Land: A state-of-the-art global reanalysis dataset for land applications, *Earth System Science Data*, 13, 4349–4383, 2021.
- 260 Nearing, G., Cohen, D., Dube, V., Gauch, M., Gilon, O., Harrigan, S., Hassidim, A., Klotz, D., Kratzert, F., Metzger, A., Nevo, S., Pappenberger, F., Prudhomme, C., Shalev, G., Shenzi, S., Tekalign, T. Y., Weitzner, D., and Matias, Y.: Global prediction of extreme floods in ungauged watersheds, *Nature*, <https://doi.org/10.1038/s41586-024-07145-1>, 2024.
- Newman, A. J., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., Viger, R. J., Blodgett, D., Brekke, L., Arnold, J. R., Hopson, T., and Duan, Q.: Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: data set characteristics and assessment of regional variability in hydrologic model performance, *Hydrology and Earth System Sciences*, 19, 209–223, <https://doi.org/10.5194/hess-19-209-2015>, 2015.
- 265 Singer, M. B., Asfaw, D. T., Rosolem, R., Cuthbert, M. O., Miralles, D. G., MacLeod, D., Quichimbo, E. A., and Michaelides, K.: Hourly potential evapotranspiration at 0.1° resolution for the global land surface from 1981-present, *Scientific Data*, <https://doi.org/10.1038/s41597-021-01003-9>.
- 270 Xie, P., Yatagai, A., Chen, M., Hayasaka, T., Fukushima, Y., Liu, C., and Yang, S.: A gauge-based analysis of daily precipitation over East Asia, *Hydrometeorol*, <https://doi.org/10.1038/s41586-024-07145-1>, 2007.