
Credit: NooPaew

REAL-WATER WATER RESOURCES MANAGEMENT: BASELINE LAND DATA ASSIMILATION SYSTEM FOR PENINSULAR INDIA

DISCLAIMER: This report builds on research initiated under the Rural Evidence and Learning for Water (REAL-Water) project, which was supported by a cooperative agreement between the United States Agency for International Development (USAID) and The Aquaya Institute. The research was completed in collaboration with Johns Hopkins University. The contents of this report are the sole responsibility of The Aquaya Institute and Johns Hopkins University and do not necessarily reflect the views of USAID or the United States Government.

ACKNOWLEDGEMENTS

The research brief was written by Benjamin Zaitchik and Sakshi Labhane, Johns Hopkins University.

PREFERRED CITATION:

REAL-Water. (2024). REAL-Water Water Resources Management: Baseline Land Data Assimilation System for Peninsular India. Rural Evidence and Learning for Water (REAL-Water) Project. This report builds on research conducted by the Rural Evidence and Learning for Water (REAL-Water) project under United States Agency for International Development (USAID) Cooperative Agreement Number 7200AA21CA00014. This report was completed independently of USAID.

AQUAYA CONTACTS:

Ranjiv Khush, Project Director
ranjiv@aquaya.org

Jeff Albert, Deputy Project Director
jeff@aquaya.org

Dayna Hansberger, Senior Program and Operations Manager
dayna@aquaya.org

ABOUT REAL-WATER:

Rural Evidence and Learning for Water (REAL-Water) was a USAID-funded applied research program that studied how to achieve safer and more sustainable rural water supply in low- and middle-income countries. Designed and originally executed as a five-year program (September 2021–September 2026) led by Aquaya, REAL-Water was terminated in February 2025 along with the vast majority of USAID's overseas development assistance programs. For further information about this and other aspects of the project, as well as to access our knowledge products, please visit https://aquaya.org/real-water-resource-hub/.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

BCSD	Bias Correction and Statistical Disaggregation
CHIRPS	Climate Hazards Group InfraRed Precipitation with Stations
CMIP	Coupled Model Intercomparison Project
ESA-CCI	European Space Agency Climate Change Initiative
ET	Evapotranspiration
FLDAS	Famine Early Warning System Land Data Assimilation System
GCM	General Circulation Model
GDAS	Global Data Assimilation System
GLDAS	Global Land Data Assimilation System
GPM	Global Precipitation Measurement
GRACE	Gravity Recovery And Climate Experiment
GSFC	Goddard Space Flight Center'
IMERG	Integrated Multi-satellite Retrievals for Global Precipitation Measurement
JHU	Johns Hopkins University
JPL	Jet Propulsion Laboratory
LDAS	Land Data Assimilation System
LIS	Land Information System
LSM	Land Surface Model
HyMAP	Hydrological Modeling and Analysis Platform
MERRA	Modern-Era Retrospective analysis for Research and Applications
MODIS	Moderate Resolution Imaging Spectroradiometer
MP	Multi-Parameterization
NASA	National Aeronautics and Space Administration
NCA-LDAS	National Climate Assessment Land Data Assimilation System
NLDAS	North American Land Data Assimilation System
SPEI	Standardised Precipitation-Evapotranspiration Index
TWS	Terrestrial Water Storage
USAID	US Agency for International Development (USAID)
WASH	Water, Sanitation, and Hygiene
WMO	World Meteorological Organization
WRM	Water Resource Management

EXECUTIVE SUMMARY

The Johns Hopkins University (JHU) research team joined REAL-Water in 2023 to support the water resources management research stream. Specifically, JHU took on responsibility for water resource modeling activities in India, which REAL-Water had identified as important for understanding the relationships between large-scale climate trends, evolving land use and water management strategies, and the sustainability of rural drinking water expansion initiatives under the Jal Jeevan Mission. Building from previous REAL-Water research and outreach on this topic, JHU implemented a modeling system for Peninsular India, focusing on the Cauvery, Krishna, and Godavari River basins (Figure 1).

In JHU Project Year 1 (REAL-Water Project Year 3), we implemented a baseline modeling system, the REAL-Water India Land Data Assimilation System (LDAS), which is described in detail below. The LDAS makes use of an advanced Land Surface Model (LSM), which generates gridded estimates of water states and fluxes across Peninsular India, a river routing model that converts water fluxes into streamflow estimates, and numerous satellite-derived datasets that serve as model parameters (e.g., vegetation and land use conditions), meteorological inputs (e.g., rainfall), and evaluation datasets (e.g., soil moisture variability). The LDAS makes use of best-available global input datasets and has been evaluated using in situ meteorological and hydrological observations and satellite imagery to confirm realism at regional scale. We find that the models capture climatic gradients, seasonality, and interannual variability in hydrological conditions across the region, albeit with larger uncertainties at local scale and in highly managed landscapes. Regional-scale realism includes general agreement with a drought inventory recently produced by leading Indian experts and consistency between model output and satellite-derived hydrological estimates. The LDAS outputs are currently stored at JHU and can be shared with partners or disseminated more broadly.

We emphasize that the REAL-Water India LDAS and the retrospective simulations generated using the system are not yet optimized for local conditions or drinking water applications, and that the simulations could also be improved with better data on soil properties and land and water management. In the coming project year, we plan to implement coupling with a more detailed groundwater model that will enable analysis of local groundwater resource conditions. We are also activating an irrigation scheme to simulate agricultural impacts on groundwater resources and selecting future climate and management scenarios (e.g., increased irrigation efficiency) that we will use to simulate potential future water resource conditions. In all of this work, we anticipate close collaboration with India-based collaborators to ensure proper use of in situ data sources and effective characterization and communication of model strengths and uncertainties.

Figure 1: River basins in India that serve as the primary focus regions for JHU modeling activities.

BACKGROUND

According to the original work plan, JHU is tasked with **generating plausible scenarios of future water availability** in the Cauvery River basin, based on recent climate variability and best estimates of future global climate change effects on regional precipitation and temperature. Conversations during Project Year 1 have led us to expand our scope to include multiple river basins in Peninsular India and to consider water resource changes associated with land use and water management in addition to climate trends. The geographic expansion is intended to maximize opportunities for model evaluation, application, and collaboration. It also decouples, to some extent, this modeling effort from politically controversial water resource policies focused on transboundary flows in the Cauvery River basin. Our focus on sustainability of rural drinking water expansion is only peripherally related to transboundary Cauvery River issues, so there is no reason to focus solely on such a hotly contested basin. The extension to consider water management and land use scenarios is a recognition that trends in water use may have a dominant impact on drinking water source sustainability in coming decades.

GOAL AND OBJECTIVES

For the first project year, our primary goal was to **implement a credible system for retrospective hydrological simulations of Peninsular India.** In doing so, we are aware that we are by no means the first group to attempt this, and that other research initiatives have produced valuable hydrological modeling systems and outputs in this region. Indeed, comparisons with these previous efforts are an important component of our workflow. The purposes of producing our own system are three-fold: (1) to provide REAL-Water with a system that the project has full control of and that can be applied to address the project's specific research needs; (2) to create an open access modeling system and simulation outputs that serve as a platform for collaborative discussion of water resource futures; (3) to leverage evolving modeling and data assimilation capabilities that may not have been available for previous modeling efforts in the region.

To this end, JHU Year 1 objectives were to:

1. Define the geographic scope of the modeling effort and identify target focus basins and geographic units.
2. Implement a baseline system for retrospective simulations that cover the period 2000-present.
3. Compare meteorological input options for the retrospective simulations.
4. Evaluate the baseline retrospective simulations to establish credibility at regional scale.

GROUNDWATER IN PENINSULAR INDIA

In taking on the challenge of simulating water resources for rural drinking water expansion in Peninsular India, we recognize that groundwater is the central resource of concern. Groundwater systems of Peninsular India are primarily fractured-crystalline cratonic rock aquifers, which make them heterogeneous and hydrologically complex. On average, these aquifers are low permeability and have lower recharge rates, as compared to unconsolidated sedimentary aquifers of northern India (Figure 2), but fracturing means that permeability and recharge potential can vary dramatically at small spatial scales. This geology presents a challenge for hydrological modeling since average aquifer characteristics are not indicative of the local conditions. It also means that aquifers of Peninsular India tend not to show the dramatic depletion trends observed in the Indo-Gangetic Plain. The recharge rate, and thus the storage capacity and sustainable yield, of these crystalline aquifers is relatively limited in the spatial average, so depletion of economically viable groundwater resources does not necessarily present as a dramatic signal in regional trends.

It is important, therefore, to interpret and apply hydrological modeling results at a spatial scale appropriate to model structure. A regional hydrological model that achieves strong performance at the scale of the aquifer or river basin is not necessarily a reliable indicator of local source sustainability. However, this doesn't mean that a regional scale model is not useful. Regional analysis can set the bounds on the water balance, offering an estimate of the constraints on total water resource development. Such models can also provide perspective on supply and demand, including spatio-temporal estimates of competing demands between agriculture and other potential uses, and they can be used to study hydrological response to climate variability and trends.

In the context of the JHU modeling effort for REAL-Water, the baseline system should be understood as a regional water balance analysis. We have not yet coupled a detailed groundwater hydrology model to capture lateral flows or fine-scale aquifer characteristics. In the next phase of the modeling effort we will incorporate those capabilities, but even then there are realistic limitations on what can be achieved when modeling fractured aquifers at scale: local flow through fractured aquifers is sensitive to site-specific fracture patterns that cannot feasibly be mapped for the entire region with existing technologies. The model should be capable of characterizing groundwater depletion risk at the scale of coherent aquifer units (i.e., continuous layers and stratigraphic units), but in the absence of detailed site-specific geophysical survey and parameterization of dedicated fine-scale groundwater flow models, we do not anticipate that we will produce a system that matches observations at single groundwater observation sites. "Source sustainability" in the context of this modeling system, then, refers to an assessment of trends in groundwater risk under changes in climate and in consumptive water use patterns, rather than predictions of whether any particular well will run dry.

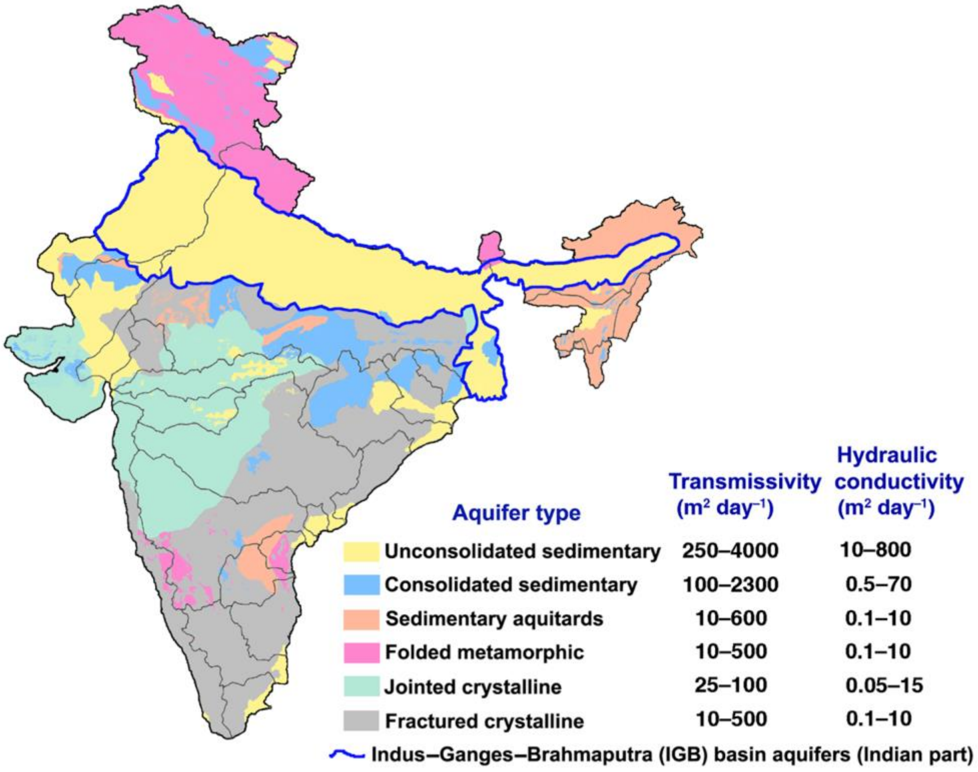

Figure 2: Major aquifer units of India, and their estimated ranges of transmissivity and hydraulic conductivity. Lower values indicate lower recharge potential (https://doi.org/10.5194/hess-23-711-2019).

LAND DATA ASSIMILATION SYSTEMS

To generate physics and observation-informed estimates of hydrological states and fluxes across Peninsular India, we implement a customized Land Data Assimilation System (LDAS). An LDAS integrates physically-based models with observational datasets – from satellite-derived products and other sources – in order to produce gridded estimates of water and energy states and fluxes, typically at spatial resolution of ~1 km^2 or coarser, that are physically consistent and spatially and temporally continuous (Figure 3). The integration of diverse water cycle observations with advanced numerical models offers a substantial advantage over using these data-streams alone. Models, when unconstrained by observations, can drift from hydrological reality. Observations, on the other hand, are incomplete representations of the water balance and have their own errors and limitations. Together, the model serves to merge observational data-streams and reconcile errors by enforcing physical consistency, while the observations keep model simulations grounded in the real world.

LDAS are commonly used in global and regional studies. Leading examples include the Global LDAS (GLDAS), North American LDAS (NLDAS) and Famine Early Warning System LDAS (FLDAS). Multiple LDAS exist, often for overlapping geographic domains, because it is important to design and optimize an LDAS for its intended use. FLDAS, for example, prioritizes data fidelity in food insecure regions and the ability to deliver results operationally with only a minimal time lag. NLDAS holds to an even tighter latency criterion on account of its more diverse end-users, but because of this it has been complemented by LDAS like the National Climate Assessment LDAS (NCA-LDAS), which seek to use a broader array of observations to reconstruct hydrological conditions of recent decades without the constraint of real-time operations. For the REAL-Water India LDAS we prioritize input datasets that have demonstrated strong performance in India, and we focus on data quality rather than latency requirements. We are also specifically concerned with trends in water storage, and for this reason we attempt to use datasets that are consistently available over the period of analysis. LDAS systems that switch rapidly between best-available input datasets tend to provide strong resource monitoring results, but they do not reliably preserve trends.

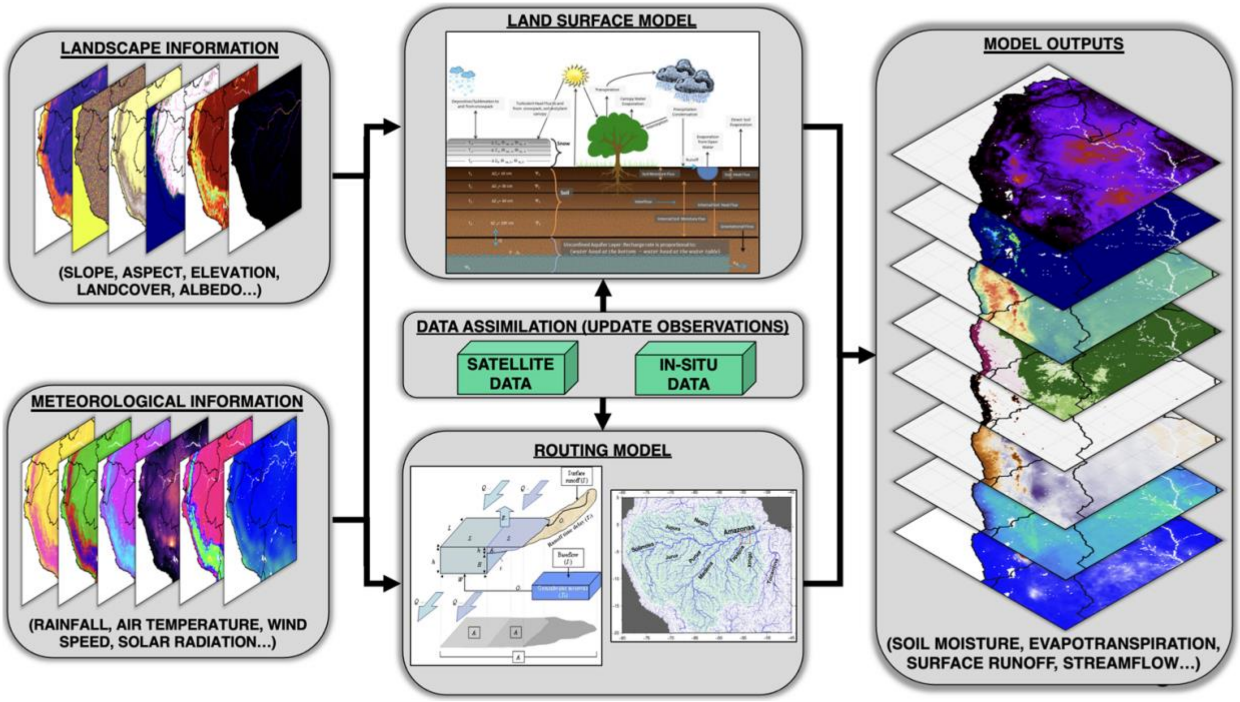

Figure 3: Schematic representation of a Land Data Assimilation System (LDAS), in which landscape and meteorological input data drive a Land Surface Model and coupled River Routing Model, both of which can be updated by the assimilation of observations of model-predicted states. The resulting gridded model outputs are products of both the models and the observations used in input and assimilation.

EVALUATION STRATEGY

The availability and quality of water resource observations in India have been a strong theme of REAL-Water project discussion in the past project year. On one hand, there is a narrative that India lacks adequate hydrometeorological and groundwater observation networks. This may be true relative to the United States and European countries. On the other hand, there are large, government-supported observational databases. Several of these holdings are free and are openly accessible within India, though they are not always accessible outside of the country. Furthermore, gaining access to large volumes of data can be practically difficult even for researchers in the country. Perhaps more importantly, the quality of government datasets is unclear. Independent researchers express significant misgivings about data quality, to the point that many choose not to use these datasets. This runs counter to the government position, expressed at the REAL-Water State of the Science workshop, held in New Delhi in July 2024, that there is ample data to support water resource analysis, and what is lacking is analytical capacity.

Questions about data quality are both a challenge and an opportunity for the LDAS. The challenge is obvious: it's not possible to calibrate or evaluate a modeling system using unreliable data. The opportunity is that the REAL-Water LDAS effort could contribute an independent perspective to the ongoing dialogue about water resource data reliability. The LDAS is, itself, full of uncertainties, and the satellite data that are used as inputs are not fully vetted for the Indian context. But the system is not intentionally biased, as some claim to be true for certain in situ datasets, and it is almost always clear what the system is simulating. For example, an LDAS simulation might not capture the water table dynamics of a given aquifer accurately, but model structure means that we know the characteristics of the aquifer and the state of human modification of the water balance in that aquifer in the model's world. When we compare LDAS results with observations, then, the presence of unusually large errors at some sites can be interpreted as a localized model error or as a sign that something is wrong with the measurement site. The provenance of all LDAS predictions is traceable to input data and parameters. Independent satellite-derived data can also be introduced to assess model realism. This then allows us to assess the characteristics of in situ data in the context of model physics and diverse remotely observable variables. Results of such analysis can help us to confirm confidence in, or doubts about, available in situ data.

Given the controversies over in situ data quality, and the need for a robust comparison of error characteristics described above, in this first-year report we focus on comparisons with globally accessible and satellite-derived datasets. The resulting evaluations give us a sense of LDAS realism. Do the LDAS Input datasets generally capture prevailing climate conditions and spatial trends? Is the seasonality of meteorology and hydrology in the LDAS credible? Are temporal trends in simulated hydrological variables consistent with those derived from independent satellite observations? They also allow us to understand some of our regional-scale uncertainties, and to pose questions about the relevance of such uncertainties to REAL-Water objectives. For example, if we drive LDAS simulations using two different satellite-derived precipitation datasets, and these simulations yield different groundwater table estimates in a given region but produce similar trends, what is the relevant result for a REAL-Water drinking water sustainability analysis? Finally, the evaluations allow us to assess where the model appears to be performing most poorly, and how we might focus our model optimization and data collection efforts to improve that performance.

REAL-WATER INDIA LDAS

REAL-Water India LDAS is implemented using the NASA Land Information System (LIS), a software framework for high performance terrestrial hydrology modeling and data assimilation that was developed to integrate satellite and ground-based observational data products and advanced modeling techniques. LIS supports multiple advanced land surface, biogeochemical, and hydrological models. In the REAL-Water India LDAS, we use the Noah MultiParameterization (Noah-MP) LSM and the Hydrological Modeling and Analysis Platform (HyMAP) river routing tool. In Project Year 1 we implemented the system at 5km resolution, with the intent to test higher resolutions (500m to 1km) for test watersheds in future project years in order to meet the source sustainability analysis capabilities identified by some partners at the 2024 State of the Science workshop.

Meteorological input data, including air temperature, specific humidity, wind speed, and downward shortwave and longwave radiation, was drawn from the MERRA-2 (Modern-Era Retrospective analysis for Research and Applications, Version 2) reanalysis product. We prepared a 20-year spin-up using NASA's Integrated Multi-satellite Retrievals for Global Precipitation Measurement (IMERG, G. Huffman et al 2015) to set up our initial conditions. We then tested two precipitation products for our final baseline simulations using (1) IMERG and (2) Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS, Funk et al 2015). Both products are satellite-based, gauge corrected products that are widely used in India and around the world. We compare the two because precipitation is such a critical input for hydrological simulation, and it is important to understand how differences between precipitation datasets propagate through the simulation of hydrological states and fluxes. Vegetation and land data information were drawn from MODIS satellite products. Soil data were drawn from the International Soil Reference and Information Centre database.

The spatial extent of simulations is 8E to 23E and 72N to 85N (260 x 300 grid cells). Linear temporal interpolation was done for meteorological forcing, and spatial interpolation was done using neighbor and bilinear methods for MERRA2 and CHIRPS/IMERG, respectively. MERRA-2 Meteorological variables were topographically downscaled according to standard lapse rate corrections.

BASELINE LDAS RESULTS

REGIONAL OVERVIEW AND SENSITIVITY TO RAINFALL DATASET

In the macro view, the meteorological data used to drive LDAS simulations capture the distribution of prevailing climate conditions across Peninsular India (Figure 4). The highest temperatures are found in low-lying areas along the coast and in interior river valleys, with cooler prevailing temperatures found at higher elevations. Variability in temperature is larger in the north, where there is strong seasonality, and lower on the coasts and in the southern peninsula, where conditions are relatively warm year-round. Rainfall is highest on the windward slopes of the Western Ghats and is lowest in the interior, with the magnitude of variability generally aligning with the patterns of total rainfall. LDAS estimated evapotranspiration is highest in the areas with high total rainfall, while variability in evapotranspiration is greatest in the interior, where seasonal dryness and episodic drought lead to greater seasonal and interannual variability than is seen in the most humid parts of the domain.

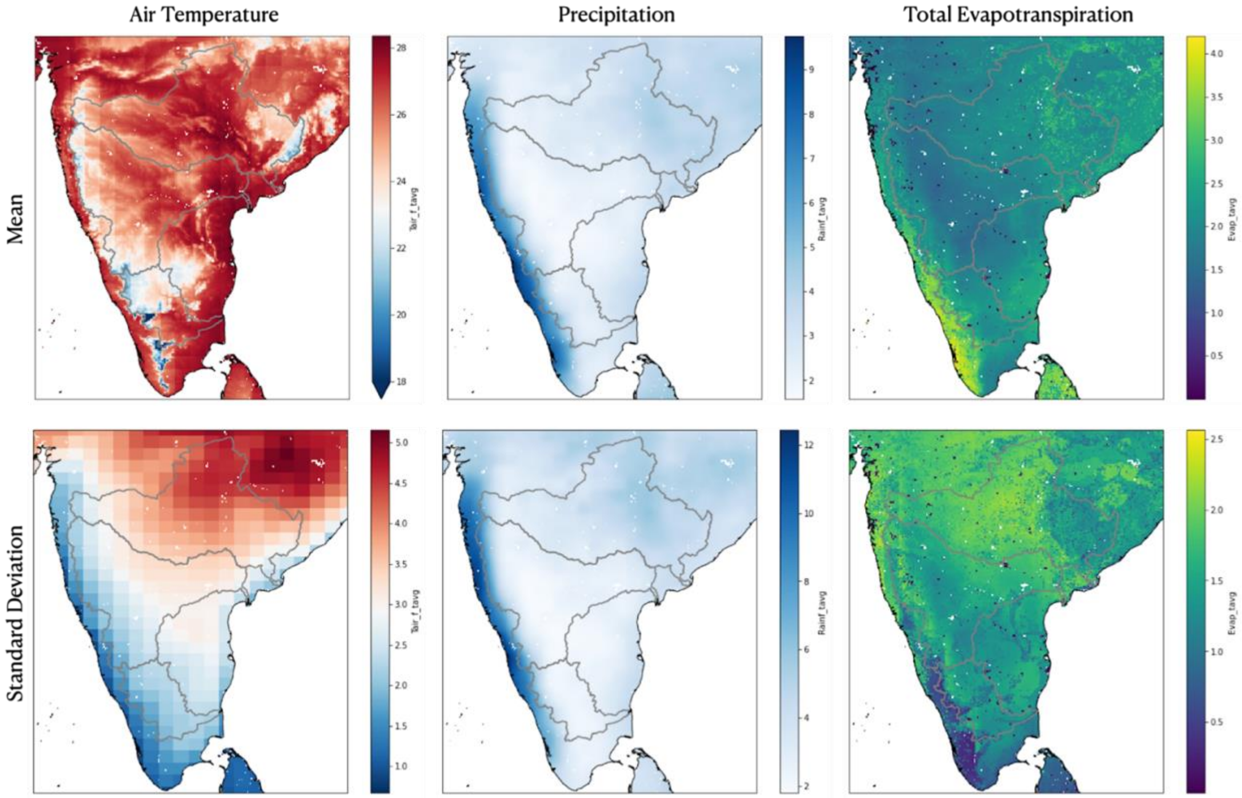

Figure 4: 2001-2020 mean temperature (degrees C), mean daily precipitation (mm/day), and mean daily evapotranspiration (mm/day), as well as monthly standard deviations of those quantities. The standard deviation indicates the magnitude of monthly variability over the period of analysis, and is larger in areas with more pronounced seasonal cycles and/or greater interannual variability. The blockier pixelated appearance of temperature standard deviation is due to the nature of temporal variability represented in GDAS, which is not as resolved as our downscaled estimates of mean temperature.

Temporally, the satellite-derived precipitation products used in the REAL-Water India LDAS show that the wettest months in Peninsular India occur during the South Asian summer monsoon, with maximum rainfall seen in August and September for Godavari and Krishna basins (Figure 5, top row). Post-monsoon rains of October and November are also appreciable, particularly in southern river basins like the Cauvery, where November is the rainiest month, on average (Figure 5, top row). The two sets of box plots in the top row of Figure 5 represent results obtained using two different precipitation products—CHIRPS and IMERG—both of which are products that integrate satellite data with in situ observations. These differences are of critical interest for us, as precipitation is the most important input to hydrological simulation. It is clear that the two products generally agree on seasonality and the relative magnitude of rainfall in the three focus basins. There are, however, potentially important discrepancies. IMERG shows wetter and more variable late monsoon rainfall than CHIRPS in the Krishna basin, for example, and CHIRPS shows a later monsoon rainfall peak in the Godavari basin than the IMERG data.

These differences in precipitation do not necessarily have a linear relationship with other hydrological fields predicted by the model. Higher peak monsoon precipitation in IMERG for the Krishna (Figure 5, top row) does not lead to higher evapotranspiration (Figure 5, bottom row) and has only modest impact on terrestrial water storage (TWS) (Figure 5, middle row). Modestly higher late monsoon precipitation in CHIRPS for the Godavari is associated with higher September TWS in that basin but relatively little difference in evapotranspiration. In the Cauvery basin, meanwhile, the products do not show obvious discrepancies in total precipitation, but CHIRPS yields higher evapotranspiration in the LDAS simulation. Nevertheless, this difference does not lead to large differences in water storage, which is the variable that is ultimately of relevance to water resource analysis. These nonlinear relationships can result from differences in the intensity or temporal autocorrelation in the precipitation fields, which will have hydrological implications that depend on the soil, vegetation, and geological conditions of the basin and on the influences of other meteorological fields.

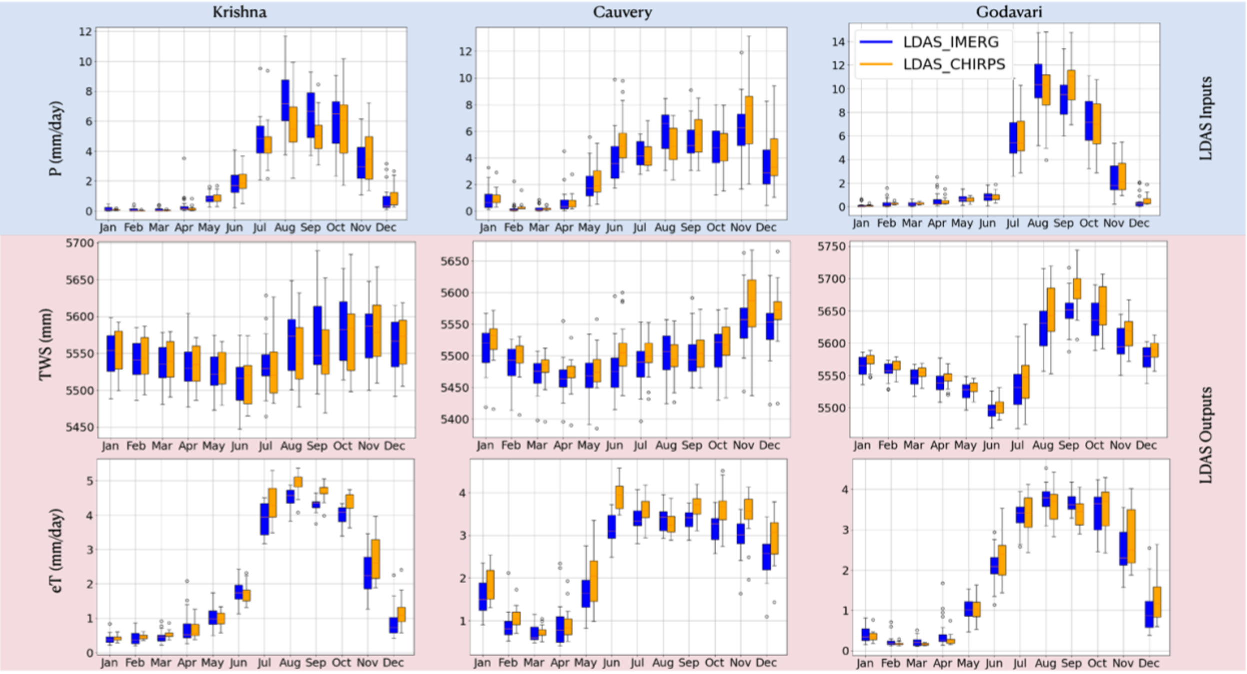

Figure 5: Boxplots showing monthly precipitation (top row), terrestrial water storage (TWS) (middle row), and evapotranspiration (ET) (bottom row). Data are for a 20-year simulation period (2001-2020) and are generated with LDAS that use either IMERG (blue) or CHIRPS (orange) precipitation data as input to the simulation.

Precipitation is shown on a blue background to indicate that it is an input, while TWS and ET are on a pink background to indicate that they are outputs. Each box shows the mean and interquartile range for the calendar month over the 20-year simulation period.

Importantly, CHIRPS and IMERG agree quite closely on interannual variability (Figure 6). Major drought years like 2015 appear clearly in both products, as do wet years like 2010. Interannual variability in precipitation propagates through to LDAS simulated hydrological fields, including runoff and TWS. As an example, Figure 7 shows IMERG-driven LDAS simulation results for the Krishna basin.

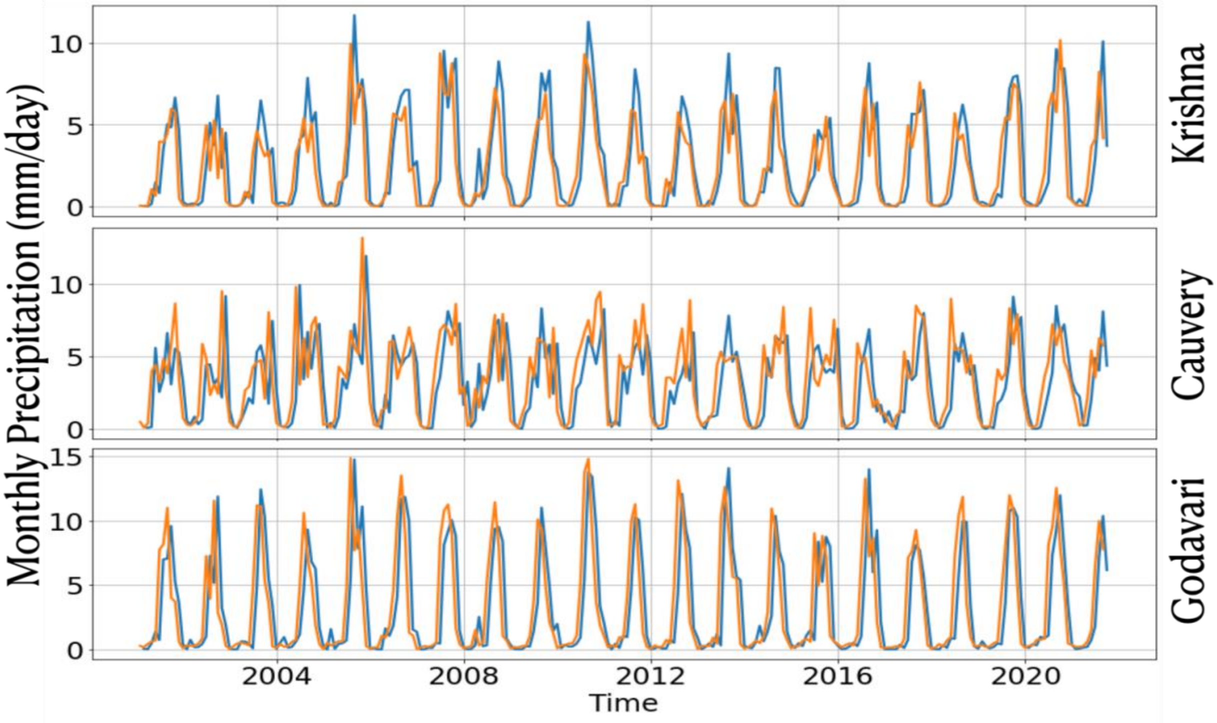

Figure 6: Monthly precipitation rates in IMERG and CHIRPS, averaged over the three target river basins.

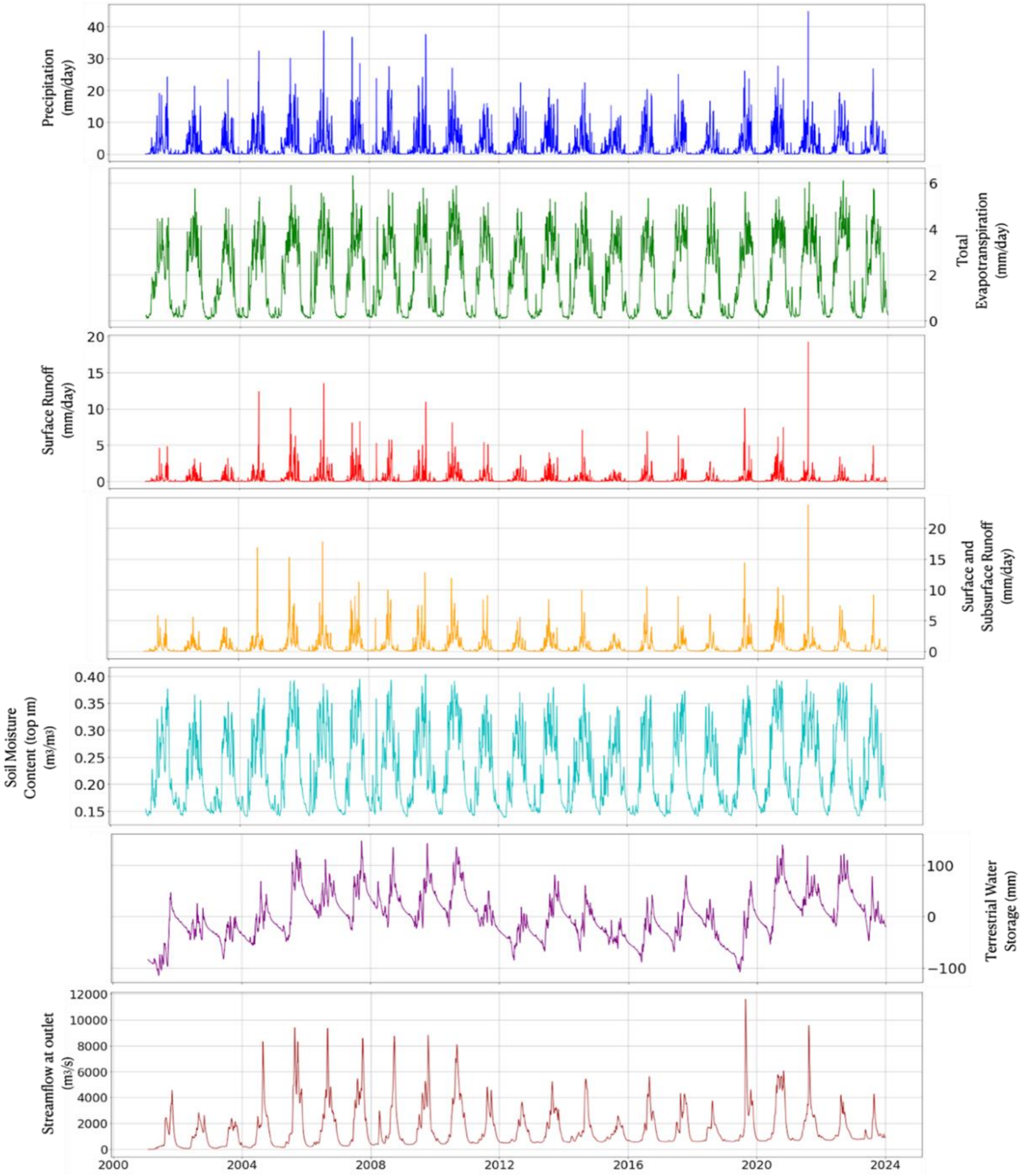

Figure 7: 3-day moving average of key variables over the Krishna Basin as represented in the LDAS using IMERG data.

The set of time series plots shown in Figure 7 capture the temporal character of hydrologic states and fluxes as represented in the LDAS. Precipitation peaks in the monsoon season but is highly variable on shorter time periods, as is expected. The land surface integrates precipitation over longer time periods, yielding evapotranspiration estimates that are smoother in time than precipitation and that exhibit a longer tail into the dry season (Figure 7, second row), as was also evident in Figure 5. Surface runoff

(Figure 7, third row), which in the LDAS refers to local storm flows that run directly from the land surface into the drainage system, is a highly variable field that shows large peaks in response to heavy rain events during the monsoon season. These are rapid response flows that occur when heavy rain falls on saturated soils, and they can be thought of as an indicator of pluvial flood risk in the modeling system. Total runoff in the LDAS, which refers to the sum of surface storm flows that run directly into the drainage system and subsurface baseflows that reach the drainage system more slowly, after passing through part of the soil column, are temporally smoother than surface runoff but are still dominated by storm water peaks (Figure 7, fourth row). We emphasize that these runoff estimates are gridded fields that represent the surface and subsurface overflow from each grid cell in the domain. They are not river discharge estimates, which are temporally smoother due to routing effects, and which are included in basin-specific results, like those shown in the bottom row of Figure 7.

Like the other fields, average soil moisture (Figure 7, fifth row) shows a rainy season peak that lags behind precipitation and evapotranspiration, as the soil column gradually dries out in the post-monsoon period. The soil moisture content in the LDAS refers to the volumetric water content of the top two meters of the soil column. Any water that infiltrates below that depth without running out of the grid cell (as a function of slope and hydraulic pressure) is accounted for in the model's shallow, unconfined aquifer. TWS anomalies, in contrast, account for variability in all water stored in vegetation, soil, and shallow groundwater in the modeling system. This field varies more gradually than soil moisture and shows more dramatic interannual variability, on account of its sensitivity to the cumulative water balance over longer time periods (Figure 7, sixth row). TWS in the baseline LDAS does not account for large-scale lateral groundwater flows, as we have not yet coupled the land surface model to an advanced groundwater model. Finally, streamflow at the basin outlet (Figure 7, bottom row) is smooth and temporally lagged relative to rainfall and runoff, as a product of the inherent spatial integration and routing processes captured by a river discharge estimate. Streamflow still responds more quickly to seasonal and interannual effects than terrestrial water storage.

Considering key features of interannual variability seen in Figure 7, the drought of 2015 is evident in all fields, as El Niño-weakened monsoon rains resulted in low runoff, reduced soil moisture, and extremely low streamflow. Soil moisture and evapotranspiration were also reduced. Interestingly, 2015 came in the midst of an extended, multi-year stretch of weaker-than-average monsoons. This is captured in the simulated TWS field, in which low values in 2015 are a combined product of weak rainfall in that year and a compromised seasonal starting point due to the relative dryness of previous years.

Wet extremes are also evident in Figure 7, including the historic Krishna River floods of 2019. The signal of that flood is more evident in the streamflow field than it is in precipitation or runoff, as it was the product of an extended period of heavy rain rather than a single day of excess precipitation that would present as a precipitation spike on these time series plots. The wet conditions of 2019 did not immediately restore TWS to historic high values, but they contributed to a recovery that took place over several years. Our ability to evaluate and more precisely simulate the groundwater dynamics that drive TWS variability will improve as we couple ParFlow into the REAL-Water India LDAS. ParFlow is an advanced groundwater model that will allow us to incorporate more detailed information on aquifer geology and account for lateral subsurface flows, enhancing the realism of simulated groundwater dynamics and allowing for higher resolution analysis of groundwater resources.

Given the relatively high agreement between CHIRPS and IMERG and the paucity of reliable observing systems that are precise enough to distinguish the marginal differences between the LDAS simulations that used either product, we have proceeded for the time being with IMERG as the primary precipitation product for the LDAS. This choice allows us to take advantage of the higher temporal resolution of IMERG and its relatively low latency, should real-time simulations become a priority for end users. CHIRPS may still be useful in some applications, however, since it has a longer record (1981-present) that can support multidecadal trend analysis. The choice of primary precipitation dataset will be revisited as needed throughout the project, preferably in consultation with end-users who may have their own opinions on the reliability of various rainfall products.

COMPARISONS WITH INDEPENDENT DATASETS

As discussed at length in stakeholder interviews and the State of the Science Workshop, the quality of in situ water resource records in India is debated, and many of our identified end users have reservations about using those records as a basis for decision-making. We believe that the in situ records, even if problematic, are an invaluable resource, and that advanced statistical techniques can be applied to sub-select reliable records and to integrate model results and in situ observations to generate optimal estimates of water resource variables. However, recognizing the controversy over these records, we limit our current data comparison to satellite-derived data and independently published water resources analyses.

One hydrological parameter that is amenable to such satellite comparisons is near-surface soil moisture, defined here as the fractional soil water content in the top ten centimeters of the soil column. This is a predicted field of Noah-MP within the REAL-Water India LDAS and is also a core satellite-derived diagnostic product of the European Space Agency's Climate Change Initiative (ESA-CCI). The ESA product is derived from passive microwave observations from multiple satellite platforms, merged in an algorithm that associates these microwave observations with near-surface volumetric soil moisture measurements collected in multiple campaigns across the globe. Like any satellite-derived product, ESA-CCI is subject to uncertainties and cannot be viewed as a gold standard value of true soil moisture. As a widely-used, spatially and temporally complete estimate of soil moisture, however, it is a valuable resource for assessing the realism of spatio-temporal variability in the LDAS.
Figure 8 shows the 20-year mean and standard deviation of monthly volumetric soil moisture in both our LDAS application (using the IMERG-driven simulation) and the ESA-CCI product. It is evident that the ESA-CCI is a spatially smoother product, reflecting its coarser resolution (25km) and its lack of information on discrete soil unit transitions. Nevertheless, general patterns hold across both products: high values in the humid areas in the southern Western Ghats, and lower values in the northeast, where soils have lower water holding capacity. Standard deviation, which primarily reflects seasonality in this analysis, is larger in the north, in response to the shorter, more seasonally peaked rainy season in that region.

To further examine temporal variability in soil moisture, we select nine districts—three in each focus basin—to look at climatology and the range of interannual variability in monthly soil moisture in the LDAS and ESA-CCI products (Figure 9). These plots reveal several key similarities between products, including their similarity in the mean, a close match in the seasonal timing of soil moisture variability, and a common identification of districts with higher peak soil moisture values (e.g., Bhandara) and lower peak soil moisture values (e.g., Raichur). There are also several key differences that require further investigation. The LDAS tends to show larger interannual variability than ESA-CCI. The LDAS shows faster drying than ESA-CCI at the end of the rainy season in some districts, and there are some differences in the minimum to which soil moisture drops in the dry season. These are differences that we will track as we add capabilities to the LDAS—in particular irrigation and lateral groundwater flow. We emphasize that these soil moisture comparisons are useful as tests of model realism and that they are relevant if the LDAS is, in the future, used by our partners for agricultural applications. Near-surface soil moisture is not directly relevant to drinking water source sustainability.

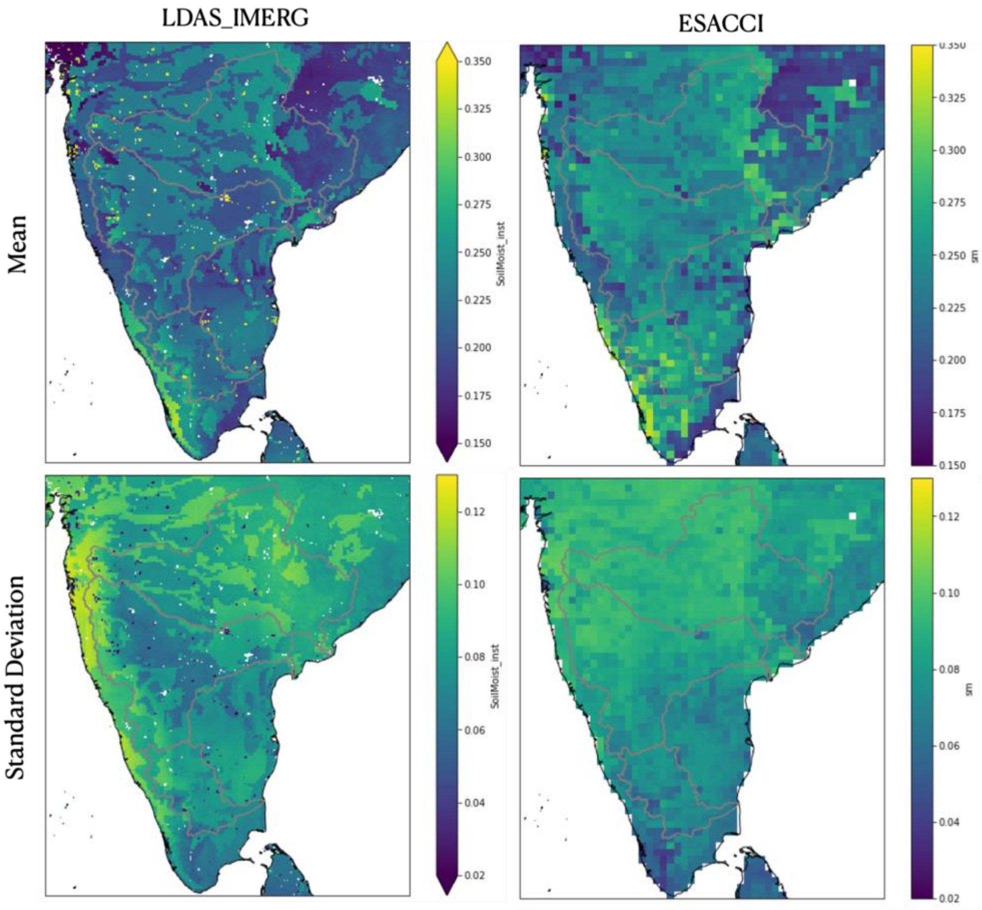

Figure 8: 20-year (2001-2020) volumetric soil moisture in the top ten centimeters of the soil column, as simulated by the LDAS (IMERG simulation) and as estimated by the ESA-CCI soil moisture product. The top row shows the mean value and the bottom row shows the standard deviation.

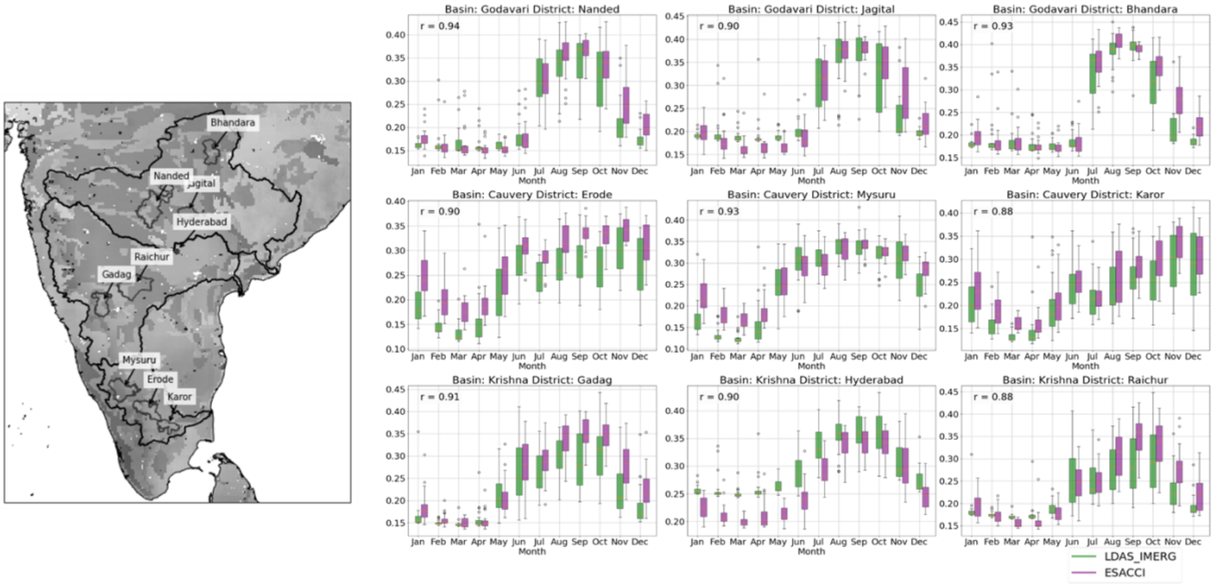

Figure 9: Districts selected for bias and seasonality check (map), and boxplots showing the seasonal cycle and interannual variability in volumetric soil moisture in the top 10 cm of the soil column in the LDAS (IMERG simulation) and ESA-CCI over the 2001-2020 period of analysis.

Another way to evaluate soil moisture variability in the model is to compare our simulated soil moisture variability to other indicators of soil dryness. The Standardized Precipitation-Evapotranspiration Index (SPEI) is a widely-used drought index that is calculated using estimates of precipitation and potential evapotranspiration. Recently, researchers led by Professor Vimal Mishra created a 1901-2020 drought atlas of India (Chuphal et al., 2024) that applies the SPEI to map the drought history of India at approximately 5km resolution. This drought atlas has its own caveats: SPEI is not a perfect proxy for soil moisture deficit and the input meteorological data are derived from sparse and sometimes error-prone records. But it is an extremely valuable reference product that was painstakingly assembled by a leading Indian research group. As such, it is a useful point of comparison for us as we seek to establish rapport with colleagues in India and establish the credibility of the REAL-Water India LDAS.

Figure 10 shows 20-year time series comparisons of the Drought Atlas SPEI and soil moisture variability in the IMERG simulation of the LDAS, performed for the same nine districts used in Figure 9. Negative values indicate drought conditions in both the LDAS and Drought Atlas time series. The magnitude of deviations are not directly comparable, as the plots compare a standardized soil moisture anomaly in LDAS to an SPEI value from the Drought Atlas, but the relative magnitude of deviations are indicative of hydrological extremes, with large negative values indicating drought. In the majority of cases, negative deviations in the LDAS correspond to SPEI drought events in all districts, and correlation values between time series are high. The character of multi-annual variability in each district is also similar across the time series. We do see isolated cases in which one of LDAS time series deviates from the Drought Atlas or the other LDAS time series—for example, the IMERG drought anomaly in Jagital in 2018. These isolated discrepancies emphasize the value of comparing multiple products and taking an ensemble approach when monitoring short-term variability. The relevance of such transient discrepancies to rural drinking water development strategy will be discussed with end-users in the next phase of LDAS modeling.

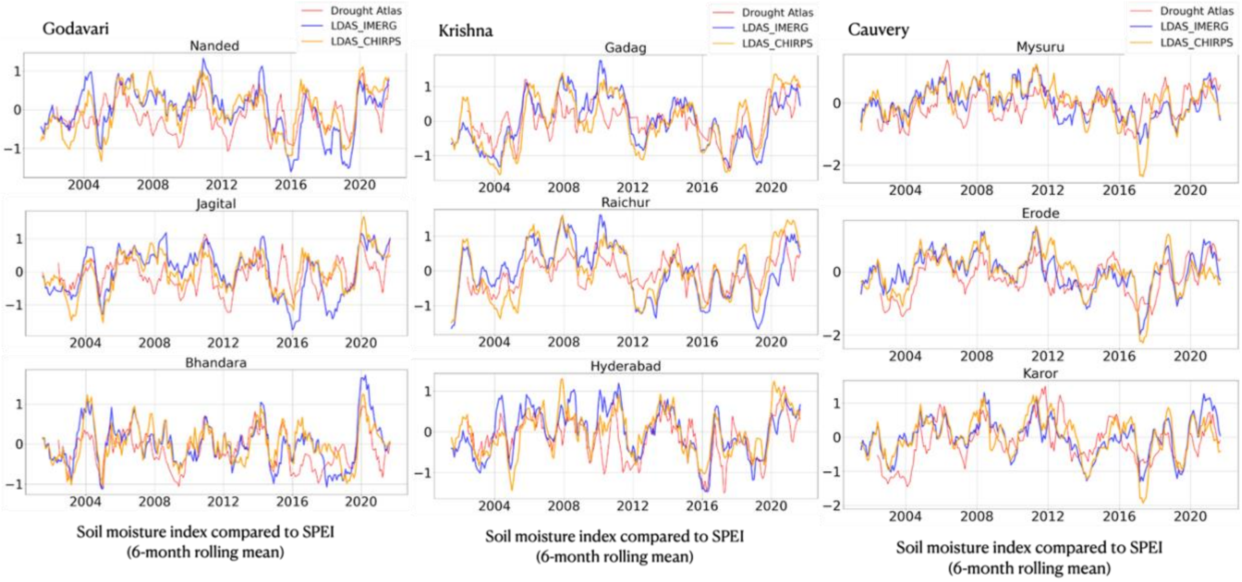

Figure 10: Time series of 6-month rolling mean of the Standardized Precipitation Evapotranspiration Index (SPEI) drawn from Chuphal et al. (2024), as compared to standardized soil moisture anomalies in the REAL-Water LDAS simulations using IMERG and CHIRPS precipitation input.

Finally, we examine LDAS performance in simulating TWS variability relative to values derived from the Gravity Recovery And Climate Experiment (GRACE) and GRACE Follow-On (GRACE-FO) satellite missions (Figures 11 & 12; for shorthand, we refer to the combined record as the "GRACE missions").

The GRACE missions derive estimated changes in water storage using measurements of change in the local gravitational field. The resulting estimates are spatially coarse—true resolution is estimated to be on the order of 300km—but offer a unique measurement of total water storage from the top of the vegetation canopy through the bottom of the deepest aquifers. These measurement characteristics have made GRACE missions an invaluable resource when monitoring variability and trends in groundwater in large, unconfined aquifers like those found in northern India, the Northeast China Plain, or the United States Great Plains. The utility of GRACE missions for source sustainability analysis in Peninsular India, where groundwater access is highly localized in fractured crystalline aquifers, is open to discussion with end users. On one hand, the GRACE mission estimates cannot be applied directly as an estimate of local source sustainability. On the other hand, they can offer a picture of regional trends and can be integrated with other data to be downscaled, to some extent, to capture some localized conditions.

Regardless of the direct utility of GRACE missions to water resource monitoring in Peninsular India, it is useful as an independent estimate of TWS that can be used when evaluating realism of the LDAS at basin scale. Figure 11 presents a time series of monthly TWS anomalies from the two LDAS simulations (IMERG and CHIRPS precipitation input) and from two GRACE-derived estimates of TWS anomaly. These two GRACE products offer one view on observation uncertainty, as they apply different mathematical solutions to the same set of GRACE mission measurements. The gap in the GRACE time series from 2017-2019 is the period between the GRACE and GRACE-FO missions. In all three focus basins we see that LDAS captures the timing of TWS seasonality and multi-annual variability in a manner similar to the GRACE missions. The magnitude of the seasonal cycle is different between GRACE missions and the LDAS. This is not surprising, given the highly integrated nature of the GRACE measurement and the fact that the baseline LDAS system uses a simplified groundwater scheme. In the next stage of modeling and end-user engagement we will investigate these differences more fully and will assess which metrics of TWS simulation are most relevant for water resources—e.g., trends, interannual variability, seasonality, total magnitude of change.

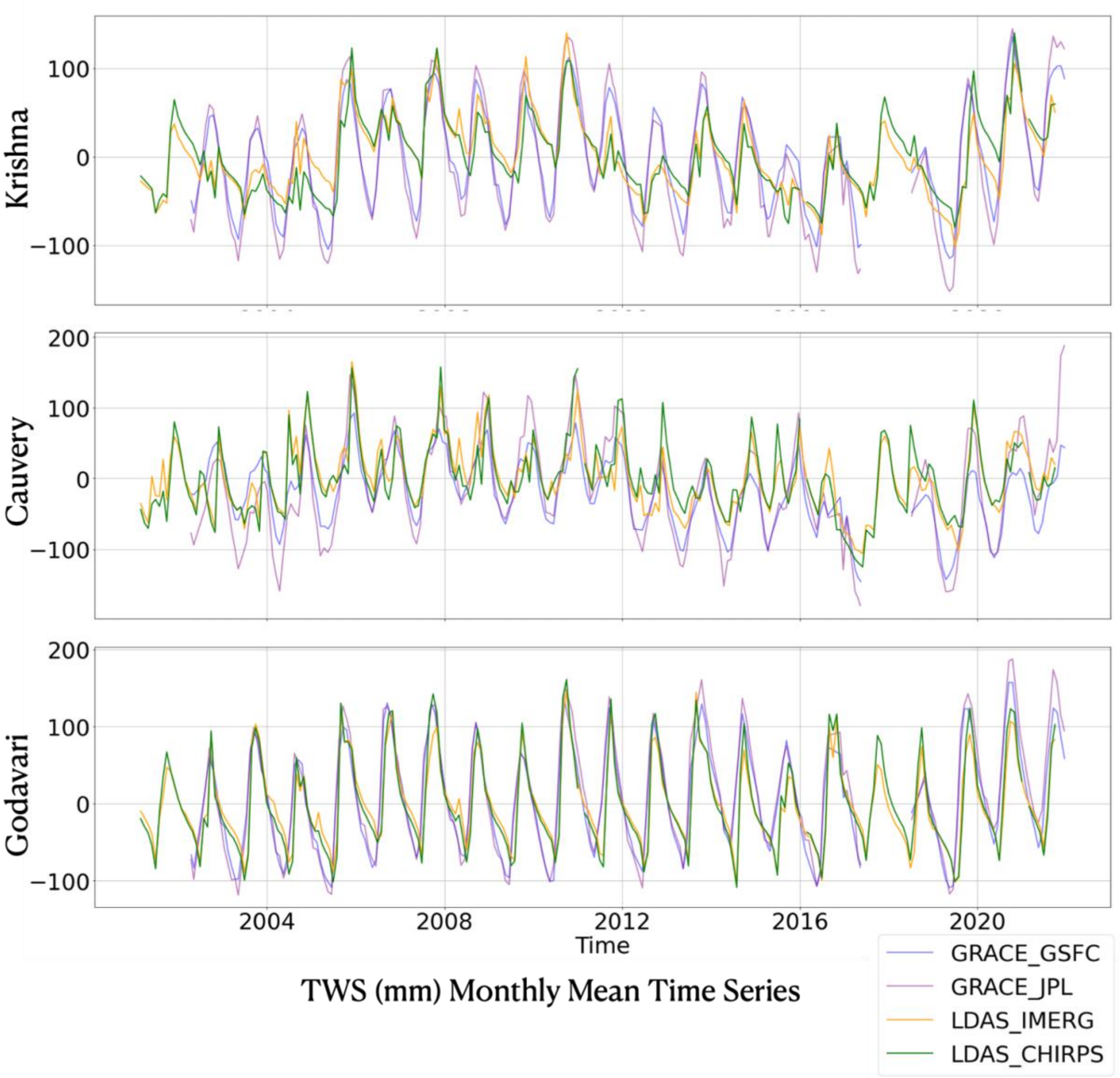

Figure 11: TWS variability according to two GRACE missions solutions, on performed by researchers at NASA Goddard Space Flight Center (GSFC) and the other by researchers at the NASA Jet Propulsion Laboratory (JPL). These products are compared to monthly TWS variability in the two LDAS simulations. Units are mm water storage for the LDAS, while the GRACE estimates are scaled by 50% to align for LDAS comparisons.

In Figure 12, we focus on interannual TWS variability in these same time series. This emphasizes the oscillation between relatively high TWS in the late 2000's, low TWS in the late 2010's, and a recent recovery in the early 2020's. That general pattern is evident in all three basins, both LDAS simulations, and both GRACE missions products. These results offer confidence that the LDAS successfully translates climatic variability into simulation of water storage variability. It does not necessarily speak to the local source sustainability of drinking water projects in these basins, given the localized nature of extractive use and trends in these aquifers. Associating these regional water storage trends with local source sustainability will require further modeling work, including downscaling with more sophisticated groundwater models (ParFlow) and adding water management components to the model.

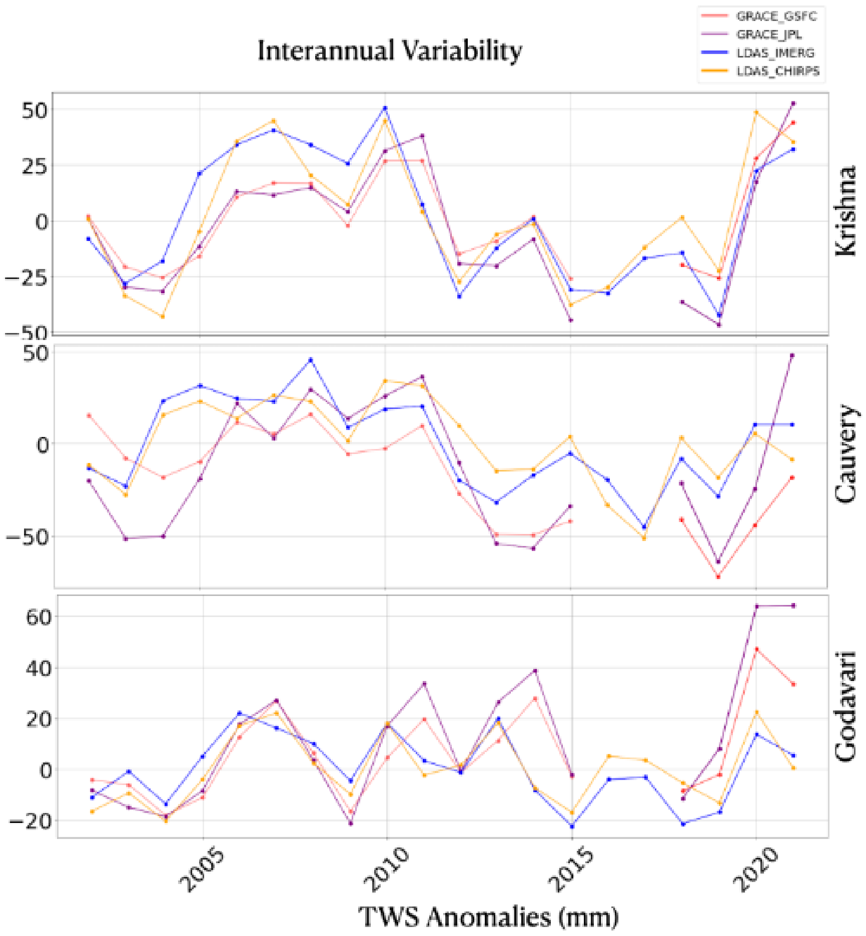

Figure 12: as in Figure 11, but showing only interannual variability in TWS.

DOWNSCALING FUTURE CLIMATE PROJECTIONS

JHU plans to apply the LDAS to projections of water balance dynamics under future climate change. This will be done by driving the surface modeling component of the LDAS with meteorological fields downscaled from the 6th Coupled Model Intercomparison Project (CMIP6) ScenarioMIP global climate model simulations. These future climate analyses are planned for the coming project year, but in Project Year 1 we established a workflow for generating downscaled meteorological fields. We have tested several downscaling methods, including multivariate regression and analog approaches, and we have found no systematic difference in performance for the India study region. For this reason, we are opting for a Bias Correction and Statistical Disaggregation (BCSD) approach (Figure 13). BCSD is an established method that we have implemented using our own code and workflow and that offers a robust approach for downscaling future climate projections from global climate models to the resolution of available observations. In our case we will target a 5km downscaled resolution, making use of GDAS and NASA GPM observations and the topographic correction routines internal to LIS. Downscaling will be applied to multiple CMIP6 GCMs and emissions scenarios, with the final choice informed by consultation with partners in India who might have strong preferences about which GCMs are selected and how the ensemble is assembled. The BCSD workflow can easily be applied to any standard GCM output. We note that the final step in the downscaling process (MetSim) is required to go from daily time steps—which is the most convenient and flexible way to work with the observation datasets—to the sub-daily temporal resolution required for reliable land surface simulation with Noah-MP. MetSim is a commonly used model for this type of temporal disaggregation.

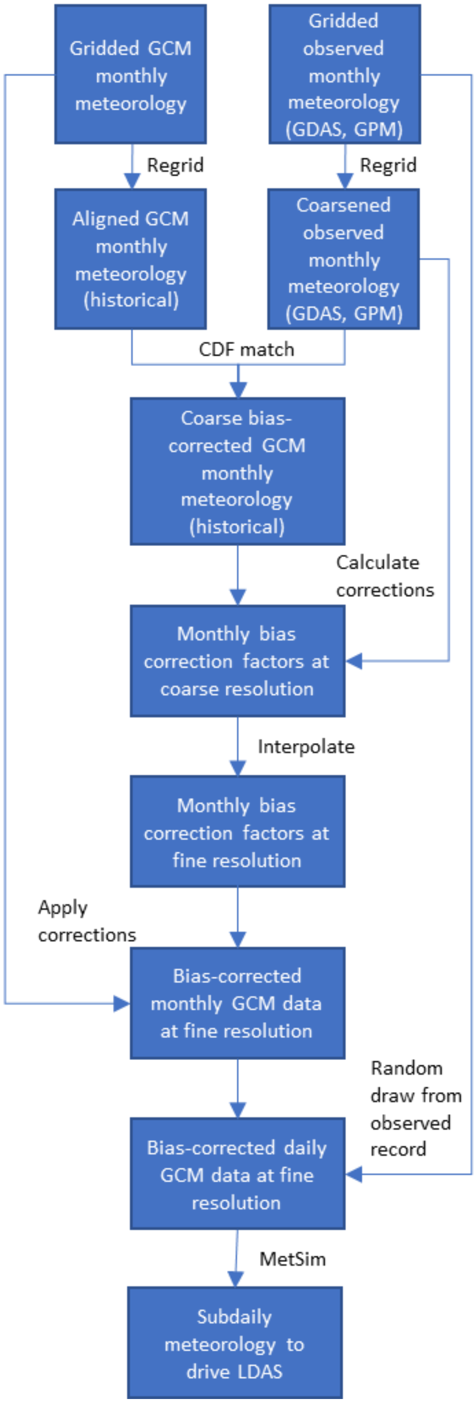

Figure 13: The Bias Correction and Spatial Disaggregation (BCSD) downscaling workflow.

NEXT STEPS

The implementation and evaluation of our baseline REAL-Water India LDAS system is now complete, and the workflow for downscaling future climate scenarios is established. As reported here, the baseline LDAS system offers credible simulation of water balance seasonality, interannual variability, and trends at regional scale, as compared to gridded estimates from merged (satellite and in situ data) observation products. We also find that the choice of precipitation dataset does not have a dominant impact on LDAS simulations of water fluxes and storages. This may not be universally true, since it would certainly be possible to derail the simulations with a highly inaccurate precipitation dataset, but IMERG and CHIRPS are both credible, widely-used products that are pinned to available World Meteorological Organization (WMO)-standard meteorological station records, and we are reassured by the fact that discrepancies between the two products do not result in large deviations in LDAS simulation.

Moving into the next phase of our work, there are several technical modeling steps required to enhance LDAS applicability to the analysis of sustainable rural drinking water expansion. These steps were anticipated in the original workplan and are already in progress. Namely, we are working to integrate representation of irrigation, groundwater withdrawal, and other direct water management activities to the model. This is critical, as rural drinking water expansion occurs on the background of massive water use and resource competition. As we consider scenarios of drinking water expansion, it is important for the project and its end-user partners to have predictive modeling capabilities that can be applied to different scenarios of water demand and management in the context of climate change. A second major modeling step is the integration of the ParFlow groundwater model to the system, as this will support simulation of groundwater hydrology at finer resolution, accounting for local aquifer heterogeneity and lateral flows.

In parallel with these model development activities, the Johns Hopkins team planned to work with the REAL-Water team to integrate the LDAS into participatory risk assessment and drinking water expansion planning. The general utility of LDAS-style modeling to water resource planning has been demonstrated in other regions and was reinforced by the India State of the Science workshop held in Project Year 3. The success of specific use cases, however, depends on the co-development of LDAS-derived data products and scenario analyses.

One key area for collaborative development of the LDAS is in the approach to water resource projections. In Project Year 4, we intended to complete the proposed future climate simulations, in which the LDAS will be used to generate hydrological predictions for the next 50 years using downscaled global climate model meteorological fields drawn from the 6th Coupled Model Intercomparison Project (CMIP6) ScenarioMIP simulations. While this offers one lens for assessing climate-related risks to water resources, participants in the State of the Science workshop emphasized their interest in projections related to water use patterns. There are multiple approaches to designing and communicating scenarios that include climate change alongside changes in water use patterns, and we will maintain flexibility as use cases are collaboratively defined in the future.

REFERENCES

Chuphal, D.S., Kushwaha, A.P., Aadhar, S. et al. Drought Atlas of India, 1901–2020. Sci Data 11, 7 (2024). https://doi.org/10.1038/s41597-023-02856-y

Dorigo, W.A., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., Chung, D., Ertl, M., Forkel, M., Gruber, A., Haas, E., Hamer, D. P. Hirschi, M., Ikonen, J., De Jeu, R. Kidd, R. Lahoz, W., Liu, Y.Y., Miralles, D., Lecomte, P. (2017). ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions. In Remote Sensing of Environment, 2017, ISSN 0034-4257, https://doi.org/10.1016/j.rse.2017.07.001

Gruber, A., Scanlon, T., van der Schalie, R., Wagner, W., and Dorigo, W. (2019). Evolution of the ESA CCI Soil Moisture climate data records and their underlying merging methodology, Earth Syst. Sci. Data, 11, 717–739, https://doi.org/10.5194/essd-11-717-2019

Preimesberger, W., Scanlon, T., Su, C. -H., Gruber, A. and Dorigo, W., "Homogenization of Structural Breaks in the Global ESA CCI Soil Moisture Multisatellite Climate Data Record," in IEEE Transactions on Geoscience and Remote Sensing, vol. 59, no. 4, pp. 2845-2862, April 2021, doi: 10.1109/TGRS.2020.3012896.