Seasonal forecasting skill for the High Mountain Asia region in the Goddard Earth Observing System

Elias C. Massoud[1,2], Lauren Andrews[3], Rolf Reichle[3], Andrea Molod[3], Jongmin Park[4], Sophie Ruehr[1], Manuela Girotto[1]

[1]University of California Berkeley, Department of Environmental Science, Policy, and Management, Berkeley, CA, USA.
[2]Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA.
[3]NASA Goddard Space Flight Center, Global Modeling & Assimilation Office, Greenbelt, MD, USA.
[4]Department of Environmental Engineering, Korea National University of Transportation, Chungju, Republic of Korea

Correspondence to: Elias C. Massoud (massoudec@ornl.gov)

Abstract. Seasonal variability of the global hydrologic cycle directly impacts human activities, including hazard assessment and mitigation, agricultural decisions, and water resources management. This is particularly true across the High Mountain Asia (HMA) region, where availability of water resources can change depending on local seasonality of the hydrologic cycle. Forecasting the atmospheric states and surface conditions, including hydrometeorological relevant variables, at subseasonal-to-seasonal (S2S) lead times of weeks-to-months is an area of active research and development. NASA's Goddard Earth Observing System (GEOS) S2S prediction system has been developed with this research goal in mind. Here, we benchmark the forecast skill of GEOS-S2S (version 2) hydrometeorological forecasts at 1-3 month lead times in the HMA region, including a portion of the Indian Subcontinent, during the retrospective forecast period, 1981-2016. To assess forecast skill, we evaluate 2-m air temperature, total precipitation, fractional snow cover, snow water equivalent, surface soil moisture, and terrestrial water storage forecasts against the Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2) and independent reanalysis data, satellite observations, and data fusion products. Anomaly correlation is highest when the forecasts are evaluated against MERRA-2 and particularly in variables with long memory in the climate system, likely due to similar initial conditions and model architecture used in GEOS-S2S and MERRA-2. When compared to MERRA-2, results for the 1-month forecast skill range from anomaly correlation of R_{anom}=0.18 for precipitation to R_{anom}=0.62 for soil moisture. Anomaly correlations are consistently lower when forecasts are evaluated against independent observations; results for the 1-month forecast skill range from R_{anom}=0.13 for snow water equivalent to R_{anom}=0.24 for fractional snow cover. We find that, generally, hydrometeorological forecast skill is dependent on the forecast lead time, the memory of the variable within the physical system, and the validation dataset used. Overall, these results benchmark the GEOS-S2S system's ability to forecast HMA hydrometeorology.

1 Introduction

Skillful prediction of hydrometeorological conditions at subseasonal-to-seasonal (S2S) timescales depends on a range of factors, including the representation of land and ocean initial conditions (Dirmeyer et al., 2018; Mariotti et al., 2018), a model's ability to capture large scale atmospheric processes (Gibson et al., 2020), a model's representation of climate mode variability (e.g., Waliser et al., 2006, 2009; Shukla et al., 2018), the chosen perturbation and ensemble scheme (Scaife et al., 2014), and the level of predictability itself. S2S forecasting differs from numerical weather prediction where skill largely depends on accurate representation of atmospheric initial conditions (Pielke Sr. et al., 1999). There is a need to understand the processes that drive S2S prediction skill, and there has been extensive research (e.g., Merryfield et al., 2020; White et al., 2021) aimed towards understanding the complexity of these systems. However, further improvements in S2S forecasting skill, particularly of societally relevant variables, are sought because accurate S2S forecasts are useful for advance planning in various sectors, such as energy, water resources, agriculture, and disaster mitigation (National Academies Press, 2016).

S2S hydrometeorological forecasts can be valuable in heavily populated regions, such as the Indian Subcontinent, as well as in more sparsely populated areas, such as High Mountain Asia (HMA). These regions experience substantial inter- and intra-annual variability in water resources. HMA has been dubbed one of the main 'water towers' of the Earth (Viviroli et al., 2007; Immerzeel et al., 2010, 2020) and has been hypothesized to influence global weather patterns through its impact on teleconnections (Nash et al., 2021). S2S forecasting systems, such as the Goddard Earth Observing System S2S prediction system (GEOS-S2S), could skillfully capture large-scale atmospheric patterns and teleconnections (Gibson et al., 2020; Lim et al., 2021), including those impacting the HMA region. Ding and Wang (2007) and Lim (2015) demonstrated the importance of the Eurasian teleconnection in driving the planetary-scale Rossby-wave propagation that causes the intraseasonal variability over central Asia and the northern part of India. Other studies investigated climate variations over HMA by the impact of the North Atlantic Oscillation (Li et al., 2005, 2008), Indian Ocean Dipole and El Niño Southern Oscillation (Stuecker et al., 2017; Sang et al., 2019; Power et al., 2021; Meena et al., 2022), the Central Indian Ocean mode (Zhou et al., 2017), and the boreal summer intraseasonal oscillation (Jiang et al., 2004; Hatsuzuka and Fujinami 2017).

The proper representation of large-scale teleconnections in S2S forecasting systems and how that impacts hydrometeorological conditions is complicated by local characteristics, which degrades the accuracy of high-resolution S2S forecasts at local scales. For example, the northward propagation of the boreal summer intraseasonal oscillation originates in the northern Indian Ocean and tends to dissipate near the foothills of Himalayas, and high humidity along the southern slope of the Himalayas and Tibetan Plateau leads to enhanced precipitation events (Jiang et al., 2004; Hatsuzuka and Fujinami 2017). It is, however, extremely difficult to pin-point specific locations where this process ultimately occurs. Therefore, to gain a better understanding and make better predictions of how the Earth system behaves at regional scales, such as for the HMA region, further research is warranted.

Numerous investigations have examined the impacts of climate and weather in the HMA region, including air temperature (Su et al., 2013; Dars et al., 2020), precipitation (Su et al., 2013; Ghatak et al., 2018; Liu and Margulis 2019; Christensen et al., 2019; Dars et al., 2020; Stanley et al., 2020), terrestrial water storage and the overall water budget (Loomis et al., 2019a; Yoon et al., 2019), groundwater storage (Xiang et al., 2016; Wang et al., 2021), snow (Liu et al., 2021a; Liu and Margulis 2019; Margulis et al., 2019), glaciers (Shugar et al., 2020; Maurer et al., 2020; Batbaatar et al., 2021), atmospheric river storms (Nash et al., 2021), hydropower (Mishra et al., 2020), and landslides (Bekaert et al., 2020; Stanley et al., 2020). There are many communities of scientists in the US, Europe, or Asia investigating how climate is changing in HMA and what drives these changes (e.g., Arendt et al., 2017).

More broadly, studies such as Vitart and Robertson (2018), de Andrade et al., (2019), and Robertson et al., (2020), have investigated the usefulness of S2S forecasting for global climate and weather extremes. These types of studies can be used to deduce the skill of S2S forecasts for the HMA region. For instance, these studies show how forecasting a variable like precipitation in the HMA region can be difficult, with forecasts being acceptable out to week 1, but starting to degrade for forecasts in weeks 2-4. Furthermore, there have been studies that investigate the skill of S2S forecasts specifically for regions including or close to HMA. For example, Deorias et al., (2021) compared the prediction of the Indian monsoon in different S2S models; Hsu et al., (2021) investigated simulations of the East Asian winter monsoon on S2S time scales; Gerlitz et al., (2020) applied climate informed seasonal forecasts of water availability in Central Asia; and Zhou et al., (2021) developed a hydrological monitoring and S2S forecasting system for south and Southeast Asian river basins. Many of these studies utilized S2S prediction systems, but there is still a need for further evaluation of S2S forecast skill for hydrometeorological variables in the HMA region. Our study examines the skill of S2S forecasting for the HMA region using the GEOS-S2S forecasting system.

The NASA Global Modeling and Assimilation Office utilizes the GEOS-S2S forecasting system, which initializes S2S forecasts each month using a weakly coupled atmosphere-ocean data assimilation system (Borovikov et al., 2018; Molod et al., 2020). Forecasts are provided to national and international multi-model prediction efforts, including the North American Multi-Model Ensemble (Kirtman et al., 2014). The skill of GEOS-S2S has been reported in various works, such as Gibson et al., (2020) who assessed the hindcast skill of representing ridging events over the Western United States in different S2S models and found the forecast horizon of GEOS-S2S to be comparable with other S2S models in the community. Recent GEOS-S2S system developments improved the representation of ocean temperatures and heat transport (Molod et al., 2020) and the retrospective forecast of climate indices, including the El Niño Southern Oscillation, North Atlantic Oscillation, and the Madden-Julian Oscillation, particularly at 1-to-3-month lead times (Molod et al., 2020; Lim et al., 2021). These improvements should contribute to enhancements in global hydrometeorological forecast skill in GEOS-S2S.

In this study, we examine the ability of GEOS-S2S forecasts to accurately predict near-surface air temperature, total precipitation, fractional snow cover area, snow water equivalent, surface soil moisture, and terrestrial water storage across the HMA region and a large portion of the Indian subcontinent at 1-, 2-, and 3-month lead times. These variables are directly relevant to the accurate prediction of water resources and processes critical to local populations. Evaluation and improvement of hydrometeorological forecast lead time can improve warning systems for natural hazards such as flooding or landslides and provide critical information for agricultural purposes (Bekaert et al., 2020; Stanley et al., 2020). However, the complex relationships among variables as well as the regional topography within HMA make S2S forecasting for this area challenging. Some of these variables, such as temperature or precipitation, are more difficult to accurately forecast at the S2S time scales compared to other variables because of their fast nature and low memory in the physical system. It is hypothesized there is higher forecast skill for the variables with longer temporal memory in the physical system, such as snow, soil moisture, or terrestrial water storage. Therefore, proper initialization of these variables can allow for longer-lasting skill in the S2S forecasting system.

The first objective of this work is to provide a benchmark of GEOS-S2S hydrometeorological forecast skill for the HMA region and across a large portion of the Indian Subcontinent. A second objective of the analysis is to determine potential areas of improvements in model initialization or more realistic representation in the model architecture, which can help enhance the forecast accuracy in future GEOS-S2S versions and extend the skillful forecast window of variables in the HMA region. The paper is organized as follows: Section 2 introduces the datasets and methods used in this study, Section 3 reports the results of the evaluation, Section 4 offers a discussion on the main findings, and Section 5 concludes with a summary of the paper.

2 Data and Methods

2.1 Region of Focus

Here, we refer to "HMA" as the domain shown in Figure 1 and covering parts of China, Afghanistan, Pakistan, Nepal, Bhutan, India, Bangladesh, Myanmar, Kazakhstan, Uzbekistan, Kyrgyzstan, and Tajikistan and stretches across several mountain ranges, including the Himalayas, Inner Tibetan Plateau, Karakoram, and Hindu Kush. These mountains funnel fresh water into major river basins that support about 1.5 billion people, providing drinking water, irrigation, and hydropower (Immerzeel et al., 2020), including the Tarim, Indus, Yangtze, and Ganges/Brahmaputra basins. HMA has one of the highest concentrations of snow and glacier ice outside of the polar regions, making it an extremely important region to study and evaluate S2S forecasting of hydrometeorological variables. The HMA region we consider here includes areas of different topography, population density, and climate. For this reason, we split our domain into different subregions (shown in the boxes in Figure 1) that account for the different areas within HMA.

2.2 GEOS-S2S Prediction System

We evaluate GEOS-S2S, version 2 (Molod et al., 2020); the GEOS-S2S forecasting system is an atmosphere-ocean general circulation model (AOGCM) and ocean data assimilation system. The AOGCM includes the GEOS atmospheric general circulation model (AGCM; Molod et al., 2015; Rienecker et al., 2008), the Catchment land surface model (Koster et al., 2000), Version 5 of the Modular Ocean Model developed by the Geophysical Fluid Dynamics Laboratory (Griffies, 2012), and Version 4.1 of the sea ice model developed by the Los Alamos National Laboratory (Hunke & Lipscomb, 2008). GEOS-S2S forecasts are initialized using a precomputed atmospheric analysis and ocean data assimilation (Penny et al., 2013). The system components are coupled using the Earth System Modeling Framework (Hill et al., 2004) and the Modeling Analysis and Prediction Layer interface layer (Suarez et al., 2007).

The GEOS-S2S analysis uses a weakly-coupled atmosphere-ocean data assimilation system with a 5-day assimilation cycle. During the initial 5-day predictor segment, every 6 hours, the departure of model trajectory from observed ocean fields is determined and sea ice fraction is replaced with satellite-derived observations (Cavalieri et al., 1996). Following the predictor segment, the model is rewound and ocean analysis increments are applied during the first 18 hours of the 5-day corrector segment. During both segments, the atmosphere is nudged to a precomputed state and SST is strongly relaxed to MERRA-2 values to ensure that the ocean and atmosphere are as consistent as possible. A detailed description is in Molod et al., (2020).

Forecasts are initialized from the GEOS-S2S analysis at the end of the corrector segment. During the retrospective forecast period (1981-2016), forecasts are initialized using an unperturbed lagged scheme, with unperturbed forecasts initialized every 5 days during the last half of each month for a total of four ensemble members. During the operational forecast period (2017-present), an additional 6 perturbed forecasts are initialized on the last forecast day of each month. All forecasts are 9 months in duration, but, we focus here on the 4 ensemble members in the retrospective forecasts with 1-, 2-, and 3-month lead times. Retrospective forecasts, which are used in this study, are completed to provide a model climatology for use in probabilistic forecasting and provide a long period for forecast verification (Molod et al., 2020). GEOS-S2S forecasts have been used and evaluated in studies related to the Madden-Julian Oscillation (Lim et al., 2021), sea surface salinity and its impact on the El Niño Southern Oscillation (Hackert et al., 2020), the impact of volcano eruptions on surface temperatures and precipitation (Aquila et al., 2021), and others.

The hydrometeorological variables of interest were obtained from the GEOS-S2S archive, and include: 2-m air temperature (T2M from the "surf" collection), total precipitation (PRECTOT from the "vis2d" collection), snow cover area fraction (ASNOW from the "vis2d" collection and called fSCA for the remainder of this paper), snow water equivalent (SNOMAS from the "vis2d" collection and called SWE for the remainder of this paper), soil moisture in the surface layer from 0-5 cm

(WET1 from the "vis2d" collection and SM for the remainder of this paper), and terrestrial water storage (TWLAND from the "surf" collection and 'TWS' for the remainder of this paper). The PRECTOT variable investigated here is total precipitation including rain and snowfall, i.e., PRECTOT = liquid + solid (total) precipitation. SM is calculated at each grid cell by scaling the WET1 variable with porosity. For grid cells that are frozen or are covered in snow, the soil moisture value is masked out as a no-data-value grid cell to focus on the warm season, following the work of De Lannoy and Reichle (2016). Simulated TWS includes soil moisture, snow, and the canopy interception reservoir, but not surface water (that is, lake and river water) or glaciers. Table 1 provides a list of these variables as represented in GEOS-S2S (Nakada et al., 2018), and the corresponding evaluation datasets, detailed in the following subsections.

2.3 Evaluation Datasets

The first product used here to evaluate the GEOS-S2S forecasts is the Modern-Era Retrospective analysis for Research and Applications, version 2 (MERRA-2; Section 2.3.1; Gelaro et al., 2017). MERRA-2 and GEOS-S2S output includes many compatible variables because the version of the GEOS AGCM in GEOS-S2S-2 is similar to the version used for the production of the MERRA-2 reanalysis.

To further evaluate GEOS-S2S, we also use independent reanalysis and observational products (Table 1). To this end, for air temperature we use the fifth-generation atmospheric reanalysis from the European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis product (ERA5; Section 2.3.2; Hersbach et al., 2020). For precipitation, we use the Asian Precipitation Highly Resolved Observed Data Integration Towards Evaluation product (APHRODITE; Section 2.3.3; Yatagai et al., 2012). For snow cover, we use the Moderate Resolution Imaging Spectroradiometer (MODIS; Section 2.3.4; Hall et al., 2002) remotely sensed product. For SWE, we use the HMA Snow Reanalysis product (HMA-SR; Section 2.3.5; Margulis et al., 2019; Liu et al., 2021b). For soil moisture, we use the European Space Agency's Climate Change Initiative data (ESA-CCI; Section 2.3.6; Dorigo et al., 2017). Lastly, for TWS, we use data from the NASA Gravity Recovery and Climate Experiment satellite mission (GRACE; Section 2.3.7; Tapley et al., 2004). We utilize information from different sources to make sure that evaluation results are not solely dependent on biases or uncertainties in a single reference product. The datasets used for evaluation in our study have their own biases and issues, particularly over the mountainous regions of our study.

2.3.1 MERRA-2

MERRA-2 is the most recent NASA global atmospheric reanalysis product and is generated using the GEOS atmospheric model and analysis (Gelaro et al. 2017). MERRA-2 output contains similar variables to GEOS-S2S and offers a rich product to apply systematic evaluation of the model forecasts. We obtain information from MERRA-2 on all the variables of interest listed in the previous section (Table 1) and for the same period (1981-2016; Bosilovich et al., 2016). The variables of interest include: 2-m air temperature (T2M from the "Single-Level Diagnostics" collection; GMAO 2015a), total precipitation

(PRECTOTCORR from the "Surface Flux Diagnostics" collection; GMAO 2015b), snow cover area fraction (FRSNO from the "Land Surface Diagnostics" collection and called fSCA for the remainder of this paper; GMAO 2015c), snow water equivalent (SNOMAS from the "Land Surface Diagnostics" collection and SWE for the remainder of this paper), soil moisture in the surface layer from 0-5 cm (GWETTOP from the "Land Surface Diagnostics" collection and SM for the remainder of this paper), and terrestrial water storage (TWLAND from the "Land Surface Diagnostics" collection and TWS for the remainder of this paper). MERRA-2 uses observation-based precipitation data as forcing for the land surface parameterization (Reichle et al., 2017), which is available as part of the "Surface Flux Diagnostics"' collection (GMAO 2015b). This PRECTOTCORR variable in MERRA-2 is compared to the PRECTOT variable in the GEOS-S2S forecasts, and similarly it contains both liquid and solid precipitation (rainfall + snowfall). Like GEOS-S2S, SM is calculated at each grid cell by multiplying the GWETTOP variable with porosity.

2.3.2 ERA5 2-m Air Temperature

For our study, we obtain information on T2M from ERA5. ERA5 covers the period from January 1950 to present and is available from the Copernicus Climate Change Service at ECMWF. ERA5 embodies a detailed record of the global atmosphere, land surface and ocean waves (Hersbach et al., 2020). The surface analysis in ERA5 ingests station observations of T2M where available and under suitable, warm-season conditions (De Rosnay et al., 2014). For times and locations where T2M observations are assimilated, the ERA5 T2M estimates are therefore closer to observations. In HMA, however, this is not necessarily the case, owing to the topographically complex terrain and generally colder conditions.

2.3.3 APHRODITE Precipitation

APHRODITE's gridded precipitation is a set of long-term, continental-scale, daily products that is based on a dense network of rain-gauge data for Asia. The data include information for many regions, including the Himalayas, South and Southeast Asia and mountainous areas in the Middle East from January 1951 until December 2015. We obtain information on the total precipitation from APHRODITE (version V1901; Yatagai et al., 2012) in our evaluation, which we utilize as the alternative observation for precipitation. The data are aggregated from daily to monthly time steps, and regridded from 0.05° to 0.5° resolution to match the model grid. There was no further quality control done on the data since this was already conducted by the data provider (Maeda et al., 2020).

2.3.4 MODIS Snow Cover Area

MODIS MOD10C1 Version 6 provides the daily (~10:30am local time) percentage of snow-covered land and cloud-covered land on the MODIS Climate Modeling Grid (posted at 0.05°; Hall and Riggs, 2016a). The MOD10C1 CMG dataset is generated from the Normalized Difference Snow Index snow cover of MOD10A1 (Hall and Riggs, 2016b) by mapping the 500 m MOD10A1 observation types (snow, snow-free land, cloud, etc.) to 0.05° bins. Snow and cloud cover percentages are derived by calculating the ratio of 500 m snow and cloud cover observations to the total number of 500 m land observations

within each CMG grid cell. MOD10C1 also has basic quality assurance flags to document any low-quality data points, and any grid cells that are flagged are not used in this analysis. Daily data between February 2000 and December 2016 are averaged to create monthly mean snow cover percentages.

2.3.5 HMA-SR Snow Water Equivalent

The HMA-SR assimilates Landsat- and MODIS-derived fractional snow-covered area to derive seasonal snow water equivalent in HMA where in situ data are limited (Margulis et al., 2019; Liu et al., 2021b). The method is a probabilistic data assimilation version of a snow reconstruction approach (Girotto et al., 2014), where SWE information is retrieved from the accumulation of melt events driven by energy forcings (i.e., downscaled global datasets for forcing a snow model) and observed snow cover area disappearance. The data product provides snow depth as well as SWE estimates from October 1999 to September 2017. The data are aggregated from daily to monthly time steps, and regridded from 16 arc-seconds (~0.0044°) to 0.5° resolution to match the model grid. We used a non-seasonal snow mask to exclude grid cells with permanent snow and ice from the evaluation (Liu et al., 2021a).

2.3.6 ESA-CCI Soil Moisture

The ESA-CCI Programme on Global Monitoring of Essential Climate Variables produces an updated soil moisture product every year (Dorigo et al., 2017; Gruber et al., 2019; Preimesberger et al., 2020). The ESA-CCI SM product comprises of active, passive, and combined microwave satellite soil moisture datasets from 1978 to 2020. In this study, information on soil moisture in the surface layer from ESA-CCI (version 6.1) is utilized. While the contributing data products represent soil moisture at varying sensing depths depending on their characteristics (active/passive sensor, measurement frequency, etc), the merged ESA-CCI SM dataset is representative of the top ~0-5 cm of the soil. There are gaps in the original ESA-CCI data due to the quality control applied during post-processing, such as areas that are masked out for ice and snow (different time steps will have different masks applied). Furthermore, the product quality changes over time with the number and type of sensors integrated into the product, with more recent retrievals being generally of higher quality.

2.3.7 GRACE Terrestrial Water Storage

From 2002 to 2017, NASA's twin Gravity Recovery and Climate Experiment (GRACE) satellites monitored large-scale water storage changes all over the globe (Tapley et al., 2004; Rodell et al., 2009; Famiglietti et al., 2011; Massoud et al., 2018, 2021, 2022). GRACE provided estimates of global mass change at monthly resolution and at a relatively coarse spatial resolution (~300 km). Information on TWS from GRACE captures the dynamic signature of all water sources on the ground, such as surface reservoirs, lakes, rivers, glaciers, canopy water, soil moisture, snow, and groundwater. For our study, we utilize the GRACE mascon product (Loomis et al., 2019b), available from April 2002 - present.

2.4 Forecast Evaluation

For the evaluation of all variables with MERRA-2, we use monthly averaged forecasts from 1981-2016. For the verification with the reference data products, we also utilize monthly averaged data, yet the time periods differ for each of the reference data products, depending on the availability and quality of the reference data. The length of record of each data product used is indicated in Table 1.

2.4.1 Calculating Climatologies and Anomalies

S2S forecasts can be assessed based on anomaly skill, i.e., the departure from expected normal conditions for a particular month (Kirtman et al., 2014). For this study, we remove the forecast climatology (i.e., the long-term mean value for each calendar month throughout the length of the available data record) for all analyzed variables. For an example of how this is estimated, consider the calculation of the 1-month lead anomaly that is initialized in January and has a forecast in February. For this, we take all the 1-month lead forecasts for February between 1981-2016 and calculate their mean. This climatology will be subtracted from the forecast of February conditions that were initialized in January (i.e., 1-month lead), to determine the anomaly for that forecast. The same procedure is applied on the 2-month and 3-month forecasts to develop their respective anomalies. For the evaluation datasets, monthly climatologies were created using the time intervals defined in the previous sections and Table 1 and subtracted from each respective dataset.

2.4.2 Regridding and Masking

All the data products listed above are remapped to a half degree resolution to match the grid size of the GEOS-S2S forecasts. For the case of higher-resolution data, such as MODIS (0.05 degree), the aggregation was done by computing the average across all the grid cells within each half degree grid cell in GEOS-S2S. For products with lower resolution, such as GRACE (1 degree posting), the data were re-gridded to half degree grids using bilinear interpolation. Furthermore, all the data is aggregated to monthly averages to facilitate the temporal comparisons with the S2S forecasts. There are cases where grid cells were excluded from the analysis for reasons such as availability or quality of the data. These excluded data were removed in the calculation of the evaluation metrics, which are described in the next section.

2.4.3 Evaluation Metrics

For evaluating the 1-, 2-, and 3-month forecasts from GEOS-S2S, we use the monthly anomalies from each data set (section 2.4.1) and estimate the unbiased Root-Mean-Square-Error (ubRMSE) as well as the anomaly correlation (R_{anom}) between the model forecasts and the reference data. For the remainder of this paper, we use ubRMSE to refer to the error and we use R_{anom} to refer to the correlation of the S2S forecasts.

The ubRMSE score is calculated as the RMSE of the anomaly forecasts from all grid cells and at all timesteps, as follows:

$$ubRMSE = 1/\sqrt{n} * \sqrt{\sum_{x,y,t} [GEOS.S2S_{anom}(x,y,t) - Ref.Data_{anom}(x,y,t)]^2}, \quad (1)$$

where $GEOS.S2S_{anom}$ is the forecast anomaly from the S2S system at location (x,y) and at time (t), $Ref.Data_{anom}$ represents the respective anomaly of the reference data product used for the evaluation, and n represents the number of elements in the calculation, which is a product of the number of grid cells and the number of time steps. For evaluation estimates that include masked-out grid cells, n is reduced to represent the total number of elements that are accounted for in the calculation.

The R_{anom} score is calculated as the correlation of the anomaly of the S2S forecasts with the anomaly of the verification data. This score is estimated using the 'corrcoef' function in MATLAB (Press et al., 1992), which also provides upper and lower limits that can be used for estimating the error bars around the correlation estimate. We report the error bars around the R_{anom} score by representing the interquartile range of the anomaly correlation from all the considered grid cells. The equation for the R_{anom} score used here is as follows:

$$R_{anom} = 1/(n-1) * \sum_{i=1}^{n} \left(\frac{GEOS.S2S_{anom}(x,y,t)}{\sigma_{GEOS}}\right)\left(\frac{Ref.Data_{anom}(x,y,t)}{\sigma_{Data}}\right), \quad (2)$$

where σ_{GEOS} and σ_{Data} are the standard deviation of the S2S forecasts and the evaluation data sets, respectively.

In this study, we also report on the ensemble spread of the GEOS-S2S forecasts. For estimating the ensemble spread for the S2S forecasts, we calculate the standard deviation of the ensemble members from GEOS-S2S for each grid cell and at each monthly time step. The ensemble spread estimated here is lead-time dependent. Since there are only 4 ensemble members at each time step and for each grid cell, the ensemble spread can be rather noisy. Therefore, we estimate the ensemble spread as the long-term mean of the standard deviation of the ensemble members at each grid cell. This helps reduce noise in the ensembles.

3 Results

In this section, we report the results of the evaluation, showing the skill of the GEOS-S2S hydrometeorological forecasts for the HMA region. For reference, Table 2 lists the ubRMSE and the R_{anom} for all variables when comparing the S2S forecasts to the reanalysis (MERRA-2, Section 2.3) and the reference data products (Section 2.4). Further discussion of the results is provided in Section 4.

3.1 Difference in Skill Among Variables and Forecast Lead Times

Table 2 and Figure 2 show the anomaly correlation for each variable and for each lead time considered, along with the error bars for each anomaly correlation assessment. The red error bars in Figure 2 indicate the spatial standard deviation of the anomaly correlation for each variable. The results indicate that across all variables, the forecast skill at 1-month lead is

higher than at 2-month lead, which is higher than at 3-month lead. For example, for T2M, the 1-month forecast anomaly correlation when compared to MERRA-2 is R_{anom}=0.24, for the 2-month forecast it is 0.13, and for the 3-month forecast it is 0.11. And when compared to ERA5, the 1-month forecast anomaly correlation for T2M is R_{anom}=0.19, for the 2-month forecast it is 0.13, and for the 3-month forecast it is 0.10. Higher anomaly correlation for forecasts with shorter lead time is seen regardless of which data product was used for the evaluation, which is aligned with other studies that evaluate S2S forecasts (e.g., Deflorio et al., 2019; Molod et al., 2020).

When comparing the S2S forecasts to MERRA-2, the variables with longer memory in the physical climate system, such as SM or TWS, have higher accuracy in the S2S forecasting system compared to variables that represent more quickly changing processes, such as T2M or PRECTOT (Figure 2a). When the reference data products are used in the evaluation (Figure 2B), results show that there is little evidence that variables with longer memory have higher forecast accuracy, since there is similar skill for forecasting most variables; for example, when evaluated against reference data the range of R_{anom} for all variables in the verification results is 0.13 to 0.24 for the 1-month lead forecasts.

Figure 3 shows the S2S forecast evaluation based on different subregions within the HMA domain. In Figure 3A-B, the ubRMSE of each box is normalized by the absolute value of the climatological mean of that climate variable in that region, then it is normalized again by all the skill values for that climate variable. For example, the ubRMSE of the West region 1-month forecast for T2M is divided by the absolute value of mean T2M in the West region (this is done to eliminate the impact of the magnitude of each climate variable in each region), then this is compared with each of the other normalized ubRMSE values of T2M for all subregions and all lead times (this is done to get a sense of which regions have more or less skill in their forecasts compared to the other regions). So in these figures, if a box is blue (red), that climate variable in that subregion for that lead time has a lower (higher) normalized error when compared to that same climate variable in other subregions and lead times. These figures show that, for example, most variables in the East region (Inner Tibetan Plateau) have a lower normalized error when compared to the other regions. Conversely, nearly all the variables in the South region (India) have a higher normalized error than other regions. Then, in Figures 3C-D, we show the original ubRMSE values for each subregion at all lead times, separated by climate variable. In these figures, errors can be compared for each region. For instance, fSCA and SWE have the highest error in the West region (Karakoram and Hindu Kush). Also, PRECTOT, SM, and TWS have the lowest error in the East (Inner Tibetan Plateau) region.

3.2 Annual Cycles

Figure 4 shows the annual cycle, averaged over HMA, of all data products considered in this study. For T2M (Figure 4A), the GEOS-S2S forecasts, MERRA-2, and ERA5 all have very similar annual cycles; this is persistent across lead times. The peak of the T2M annual cycle occurs during the summer months (June, July, August) reaching 290-295 K, and the low occurs during the winter months (December, January February) dropping to 273-275 K. GEOS-S2S PRECTOT forecasts

(Figure 4B) have a wet bias compared to the MERRA-2 and APHRODITE products across nearly the entire annual cycle. The peak of all products occurs in the summer months (JJA) reaching ~4-4.5 mm day^{-1} for the S2S forecasts and 3.5-4 mm day^{-1} for the evaluation products, and a low in the winter months (NDJF) dropping to 0.5-0.75 mm day^{-1} for the S2S forecasts and 0.25-0.5 mm day^{-1} for the evaluation products.

GEOS-S2S fSCA forecasts have more snow cover compared to the MERRA-2 and MODIS products, with a consistently higher mean and amplitude in the S2S forecasts (Figure 4C). As expected, the peak of all fSCA products occurs in the winter months (DJF) and the low in the summer months (JAS), however the magnitude and amplitude are different between the products. The peak fSCA in the S2S forecasts reaches ~0.23-0.25 and the low is about 0.05 for the S2S forecasts, and for the evaluation products the peak is only about 0.1-0.15 with lows that are less than 0.03. Furthermore, the annual cycle of fSCA in MERRA-2 is consistently the lowest out of all the products. For SWE (Figure 4D), the GEOS-S2S forecasts have a different annual mean and amplitude for the various lead times and in comparison to the evaluation products. The peak SWE in the S2S forecasts occurs in the spring months (March and April), reaching a high of about 0.02 m for the 1-month lead forecasts, 0.025 m for the 2-month lead forecasts, and 0.03 m for the 3-month lead forecasts. For the HMA-SR product, the seasonality has a higher amplitude and magnitude, reaching a peak of ~0.03 m in the spring months. For MERRA-2, the annual cycle of SWE is consistently lower compared to the other products, reaching a peak of ~0.005 m in February. All products show a minimum SWE of less than 0.005 m in the summer months.

For SM, the annual cycle of the GEOS-S2S forecasts is like that of the MERRA-2 product but is substantially different from the annual cycle of the ESA-CCI data (Figure 4E). The peak SM in the S2S forecasts occurs in the fall (~October) and reaches ~0.25 $m^3 m^{-3}$ and the low occurs in the spring (around May) and drops to about 0.12 $m^3 m^{-3}$. This is similar in MERRA-2, with a peak of just over 0.2 $m^3 m^{-3}$ that occurs in November and a low of just under 0.15 $m^3 m^{-3}$ that occurs in May. For the ESA-CCI, the peak SM reaches ~0.27 $m^3 m^{-3}$ and is observed in the summer months (JJAS) and the low drops to 0.17 $m^3 m^{-3}$ and is observed in the early spring (March). Similarly, for TWS (Figure 4D), the annual cycle of the GEOS-S2S forecasts is like that of the MERRA-2 product but is substantially different from the annual cycles of the GRACE data. The peak of the TWS anomaly in the S2S forecasts and in MERRA-2 occurs in the late summer (ASO) and reaches ~0.05 m and the low occurs in the spring (April and May) and drops to about -0.02 m. For GRACE, the peak TWS reaches a high of 0.05 m in the summer (JJA) and drops to a low of -0.05 m in the spring (March-April). Therefore, there is a 1–2-month temporal lag as well as a difference in the mean and amplitude of the annual cycles of the different products of SM and TWS between the various products considered.

3.3 Error by Forecast Month

The S2S forecast skill depends on various factors, such as the lead time or the variable of interest. Our results in Figures 5 and 6 show that skill also depends on the month that is forecasted. We observe this behavior in GEOS-S2S when compared

to both MERRA-2 (Figure 5) and reference data products (Figure 6). Note, the y-axes in Figures 5 and 6 are different so that the seasonality in each figure is properly portrayed. Figures 5 and 6 show the area-averaged error based on the forecast month of interest for each variable. As an example, the three bars for the month of April include the 1-month, 2-month, and 3-month forecasts, which are the forecasts initialized in March, February, and January respectively. These results can be suitable for those interested in understanding the errors of a forecast for a specific month, say April, using forecasts that were made 1 to 3 months prior.

We find that GEOS-S2S forecasts of T2M have less skill in the winter season with ubRMSE greater than 2 K around February and more skill in the summer season with ubRMSE of less than 1.5 K around August (Figures 5A and 6A). Errors in the precipitation forecasts are higher in the summer (July-August) compared to the winter months (December-January), with ubRMSE that is greater than 2 mm day^{-1} in the summer and less than 0.5 mm day^{-1} in the winter (Figures 5B and 6B).

For the snow variables, forecasts of fSCA have higher errors in the winter season (December-February) with ubRMSE close to 0.1, and less error in the summer season (July-August) with ubRMSE of less than 0.01 (Figures 5C and 6C). For SWE, results are different when comparing the S2S forecasts with MERRA-2 and with the HMA-SR product. Figure 5D shows that when comparing the S2S forecasts of SWE to MERRA-2, there are higher errors in the spring (March-April) with ubRMSE of 1-1.5 cm and lower errors in the summer (August-September) with ubRMSE of less than 0.1 cm, with the forecast lead time impacting the amount of error. Yet, Figure 6D shows that when comparing the S2S forecasts of SWE to the HMA-SR product, there are higher errors in the summer months (July-August) with ubRMSE of 4 cm and lower errors in the fall (October-November) with ubRMSE close to 1 cm.

Errors in the SM forecasts are higher in the summer (July-August) compared to the winter months (February-April), with ubRMSE values up to 0.03 $m^3 m^{-3}$ in the summer and as low as 0.01 $m^3 m^{-3}$ in the winter and spring (Figures 5E and 6E), and with the forecast lead time impacting the magnitude of error. For TWS forecasts, results are different when comparing the S2S forecasts with MERRA-2 and with the GRACE data. Figure 5F shows that when comparing the S2S forecasts of TWS to MERRA-2, there are higher errors in the summer (around August) with ubRMSE that is over 4 cm and lower errors in the winter (around February) with ubRMSE as low as 2 cm, with the forecast lead time impacting the magnitude of error. Yet, Figure 6F shows that when comparing the S2S forecasts of TWS to the GRACE data, there are higher errors in the spring months (around April) with ubRMSE greater than 15 cm and lower errors in the winter (around February) with ubRMSE below 10 cm.

3.4 Spatial Patterns: Climatology, Ensemble Spread, and Forecast Error

This section focuses on the spatial aspect of the evaluation. Figures 7-12 show, for each variable, the GEOS-S2S ensemble mean climatology (1981-2016), the ensemble spread, the ubRMSE versus MERRA-2, and the ubRMSE versus the reference

data products. The top rows of these figures show the results for the 1-month forecasts, and the middle and bottom rows show the differences in the 2-month and 3-month forecasts with respect to the 1-month forecasts.

3.4.1 Evaluation of Temperature and Precipitation

As expected, T2M is generally higher at lower elevations, for example in India and Pakistan, and it is much cooler in the mountains and at higher elevation, for example in the Himalayas and the Inner Tibetan Plateau (Figure 7A). The ensemble spread of T2M (Figure 7B) is low compared to the ubRMSE (Figure 7C and 7D), indicating that most ensemble members forecast similar T2M values. The ubRMSE is larger in regions where the spread is higher, indicating that the spatial patterns of the ensemble spread and ubRMSE are similar. The ubRMSE values relative to MERRA-2 and ERA5 show a similar magnitude throughout most of the domain, with ubRMSE values of up to ~3 K (Figure 7C and 7D). However, for the Inner Tibetan Plateau, there is more agreement with MERRA-2 (ubRMSE of ~2 K) than with the ERA5 product (ubRMSE of ~3 K). The 2-month (Figure 7E) and 3-month (Figure 7I) forecasts show a progressively warmer Indian subcontinent but are cooler in the remainder of the domain compared to the 1-month forecast. Furthermore, the ensemble spread (Figure 7F and 7J) and ubRMSE (Figure 7G, 7H, 7K, and 7L) generally increase with increasing lead times, except for the Pakistan region. Notably, the increase in ensemble spread and error with increasing lead time is greatest in India and less pronounced for the Tibetan Plateau. These results reinforce the findings from Figure 3 that show the evaluation based on subregions.

The mean climatology of precipitation is much wetter in parts of the domain with higher gradients of elevation, with greater than 15 mm day^{-1} in the mountain ranges (e.g., Himalayas) and less than 5 mm day^{-1} for other parts of the domain (Figure 8A). Furthermore, for these same regions, the ensemble spread of PRECTOT is also much higher compared to other parts of the domain (Figure 8B), with a mean ensemble spread up to 6 mm day^{-1}. The comparisons with MERRA-2 (Figure 8C) and APHRODITE (Figure 8D) both show a similar magnitude of error throughout most of the domain. The largest errors in PRECTOT forecasts are in the Indian subcontinent and in the Himalayas (ubRMSE up to 5 mm day^{-1}). This result matches the spatial interpretation of the subregion analysis shown in Figure 3C-D. The 2-month (Figure 8E) and 3-month (Figure 8I) forecasts show a drier Indian subcontinent and are somewhat wetter in regions with high elevation when compared to the 1-month forecast. This difference tends to propagate to other variables shown in later figures, i.e., higher fSCA and SWE values in the mountains and lower SM and TWS values in the Indian subcontinent for 2- and 3-month compared to 1-month forecasts. Furthermore, the ensemble spread is generally higher in the mountain regions and lower over the Indian subcontinent with increasing lead time (Figure 8F and 8J). For the error in precipitation, there are regions with higher error in forecasts with longer lead times, such as in India, and other regions where the error is lower with longer lead times, such as in Southeast Asia (Figure 8G, 8H, 8K, and 8L).

3.4.2 Evaluation of Snow Cover Area and Snow Water Equivalent

Snow cover is generally only found in the regions of the domain with high elevation (Figure 9A), and there is much more snow-covered area in the north-western parts of the domain (e.g., Hindu Kush and Karakoram). The ensemble spread of fSCA (Figure 9B) is high for much of the domain where there is snow cover, including the Himalayas and the Inner Tibetan Plateau. The 2-month (Figure 9E) and 3-month (Figure 9I) forecasts show higher amounts of fSCA for much of the domain compared to the 1-month forecasts (Figure 9A), which could be attributed to the fact that at longer lead times the forecasts are colder and wetter at higher elevations (Figures 7-8 panels A, E, and I). The comparison with MERRA-2 (Figure 9C) and MODIS (Figure 9D) both show that errors are present where there is snow cover, where the grid cells that have no snow cover are masked out. This result matches the spatial interpretation of the subregion analysis shown in Figure 3C-D. The error compared to MERRA-2 (ubRMSE close to 0.2) is noticeably higher than the error compared to MODIS (ubRMSE close to 0.1), especially for regions with high fSCA. This shows that GEOS-S2S fSCA is closer to what is shown in MODIS than to the MERRA-2 product, which supports the results from Figure 4C. Additionally, the ensemble spread (Figure 9F and 9J) and the forecast errors (Figure 9G, 9H, 9K, and 9L) generally increase in the mountain regions with increasing lead time.

Similarly, the mean climatology of SWE (Figure 10A) indicates that snow is present in the regions of the domain with high elevation, specifically in the major mountain ranges. Consequently, the ensemble spread of SWE (Figure 10B) is also high in these locations (mean spread up to 0.05 m) and very low elsewhere in the domain (mean spread less than 0.01 m). The 2-month (Figure 10E) and 3-month (Figure 10I) forecasts show higher amounts of SWE in the major mountain ranges, which again could be attributed to the fact that at longer lead times the forecasts are colder (Figure 7E and 7I) and wetter (Figure 8E and 8I) in regions with high elevation gradients. The ubRMSE maps vs. MERRA-2 (Figure 10C) and HMA-SR (Figure 10D) show that errors are higher where there is more snow, which is expected. Again, this result matches the spatial interpretation of the subregion analysis shown in Figure 3C-D. Here, the error compared to HMA-SR is considerably higher (ubRMSE up to 0.1 m) than the error compared to MERRA-2 (ubRMSE up to 0.04 m), especially for regions with high SWE. And like fSCA, the ensemble spread (Figure 10F and 10J) and the forecast errors (Figure 10G, 10H, 10K, and 10L) are generally higher with increasing lead times, particularly in the major mountain ranges.

3.4.1 Evaluation of Soil Moisture and Terrestrial Water Storage

The mean climatology of SM (Figure 11A) shows that soil moisture is high in India and Southeast Asia (~0.4 $m^3 m^{-3}$) and is low in the western and northern parts of the domain (~0.1 $m^3 m^{-3}$). There are lower SM values for forecasts with increasing lead times for the Indian subcontinent (Figure 11E and 11I). This could be attributed to the fact that at longer lead times the forecasts are hotter (Figure 7E and 7I) and have less precipitation (Figure 8E and 8I) across the Indian subcontinent. However, for Myanmar and Southeast Asia, longer lead times produce higher SM values. The ensemble spread of SM (Figure 11B) is lower for the 1-month forecasts and increases in magnitude for longer lead times (Figure 11F and 11J). The

ubRMSE maps vs. MERRA-2 (Figure 11C) and ESA-CCI (Figure 11D) report higher errors over regions with higher soil moisture values (ubRMSE of up to 0.06 $m^3 m^{-3}$). This result matches the spatial interpretation of the subregion analysis shown in Figure 3C-D. Furthermore, the error increases with lead time (Figure 11G, 11H, 11K, and 11L), especially in India when compared to MERRA-2 (Figure 11G and 11K). However, when compared to ESA-CCI, the forecast error decreases with lead time for the western and northern parts of the domain (figure 11H and 11L). Additional data gaps are shown in these figures due to snow covered and frozen grid cells being masked out in the S2S forecasts and due to quality control applied during post-processing of the ESA-CCI product.

The mean climatology of TWS (Figure 12A) shows that water storage is higher in Myanmar and Southeast Asia and is lower in the other parts of the domain. The ensemble spread of TWS (Figure 12B) is higher in the regions with high elevation gradient (e.g., Himalayan Mountain range). Additionally, the spread of TWS is lower for the 1-month forecasts and increases in magnitude with lead time (Figure 12F and 12J). The evaluation of GEOS-S2S forecasts of TWS show that the forecasts are much closer to MERRA-2 (Figure 12C, ubRMSE less than 0.1 m) than to GRACE (Figure 12D, ubRMSE up to 0.3 m). The errors compared to GRACE are 3-4 times higher in many regions, especially for the Indian subcontinent, Myanmar, and Southeast Asia (Figure 12D). Again, this result matches the spatial interpretation of the subregion analysis shown in Figure 3C-D. When compared to MERRA-2, forecasts with longer lead time (Figure 12G and 12K) have higher errors, yet when compared to GRACE, there is no consistent change in the error with longer lead times (Figure 12H and 12L), with some regions such as in India having less error with longer lead times.

4 Discussion

4.1 Role of Model Initialization and Hydrologic Persistence

S2S forecasting for HMA is in its infancy. Skill has historically been somewhat low and, as shown in our results, certain variables have high forecast skill while others are more difficult to forecast. When comparing the S2S forecasts with MERRA-2, Figures 2A and 3A show that the snow variables, SM, and TWS have relatively higher skill at early lead time (1-month), and for SM and TWS, this skill can persist for forecasts at longer lead time (2-3 months). This could be because GEOS-S2S and MERRA-2 have similar land conditions during initialization, both modeling systems are quite similar, and because these variables have longer persistence and memory in the physical system. When evaluating the S2S forecasts against MERRA-2, forecast skill is highest in long-memory variables (snow and soil moisture related) and lower in near surface atmospheric variables (T2M and precipitation). In all instances, forecast skill decreases rapidly with increasing forecast lead time. When comparing the S2S forecasts with reference data products (Figures 2B and 3B), the decline in forecast skill across lead times is slower and the anomaly correlations are not consistently statistically different across lead times.

Another reason that could explain the skill in certain variables is the role of better land surface initial conditions. For example, fSCA, SWE, SM, and TWS vary more slowly compared to T2M or PRECTOT, and their initial conditions play an important role in the skill of 1-month forecasts. This can be inferred in our results. For example, in Figure 2A the forecast skill relative to MERRA-2 is higher for these variables, perhaps due to similar initialization in the GEOS-S2S and MERRA-2 systems. However, in Figure 2B, the forecast skill relative to the reference data products is not as high. Furthermore, in a more regional sense, it is possible that improvements in model initialization for SWE and SM may translate to improvements in forecast skill for the West and East subregions (Karakoram and Inner Tibet Plateau); the evidence for this is shown in Figure 3AB where higher skill can be seen in the 1-month forecast for SWE and SM for these regions when evaluated against MERRA-2 compared to the evaluation against the other observations. Enhancements in forecast skill due to improved model initialization for these processes with slower temporal dynamics has been shown in other studies as well (Getirana et al., 2020; Zhou et al., 2021). Therefore, forecast skill in shorter memory variables (T2M, PRECTOT) may increase with improvements in resolution and process representation, and gains in forecast skill for longer memory variables (fSCA, SWE, SM, and TWS) may be achieved with improved land surface initial conditions, and if successful, increased forecast skill in 1-month lead time can propagate through to longer leads.

4.2 Reliability of S2S Forecasts

Other than looking at the forecast error to determine whether a forecast was skillful or not, the spread of the forecast ensemble is another metric that gives indication of reliability when preparing for impacts of weather events. For instance, a smaller spread in the S2S forecasts for a given region might be an indication of higher skill for that variable in that region. The results shown in this study, such as those in Supplementary Figures S1-S3, provide a benchmark of information regarding the forecast skill as well as the ensemble spread in the GEOS-S2S seasonal forecasts. Generally, one can compute the spread/error ratio with the goal of that being close to 1; if it is larger than 1 (more spread than error) this is considered "underconfident", and if it is less than 1 this is considered "overconfident" (Fortin et al., 2014). For the reliability plots in Figures S1-S3, almost all the maps are blue, indicating that the forecasts are overconfident, meaning there is a smaller spread compared to what the error is. However, for SM (Figure S3) this is the opposite, with red indicating that the forecasts are underconfident, meaning there is a larger spread compared to what the error is. Furthermore, for PRECTOT, fSCA, and SWE (Figures S1-S2), there are regions in the Karakoram, Himalayas, and Inner Tibetan Plateau that also show red, indicating that the forecasts are underconfident. It is important to note, however, that there are limitations to using this reliability metric, including the fact that one can have a "perfect" ensemble prediction system with low correlation between skill and spread (c.f., Hopson 2014), in which case the reliability of the forecasts would be difficult to capture.

4.3 The Role of Model Characteristics

4.3.1 Resolution

For parts of the domain with high elevation and high topographic variability, many of the variables including PRECTOT, fSCA, SWE, and TWS had large errors (Figures 8C-D, 9C, and 10D) as well as large ensemble spreads (Figures 8B, 9B, 10B, and 12B). This is an indication of the difficulties of accurately forecasting climate for regions of high elevation and complex topography. This could be because of the coarse spatial resolution of the GEOS-S2S simulations with topography posted at a 0.5-degree resolution (i.e., ~ 50 km). The topographic smoothness in the model can impact the simulations in various ways, such as limited orographic effects or issues associated with the formation and propagation of weather events. To confirm this argument, Cannon et al., (2017) discussed the effects of topographic smoothing on the simulation of winter precipitation in HMA and found that precipitation distributions in topography that is represented in experiments with coarser resolution are biased relative to a simulation with more realistic topography. Furthermore, Zhou et al., (2021) used optimized land initial conditions from GEOS-S2S, and they were able to downscale outputs of soil moisture to 5 km resolution and assess the forecast time horizon out to 9 months. Therefore, resolution can have a contribution to forecast skill, and it is possible that improved resolution in the S2S forecasts can help to enhance the forecast skill of certain variables.

4.3.2 Seasonality: Representing the Monsoon and Other Atmospheric Processes

S2S forecast skill largely depends on getting the seasonal signature in the forecasting system correctly. In our results, there are seasonal patterns in the GEOS-S2S forecast skill (Figures 4-6), and the simulated seasonality and the annual cycle of the hydrometeorological variables are generally well captured. For example, T2M and PRECTOT errors vary in relation to the Indian monsoon season (JJAS). Precipitation error tends to increase in these months (Figure 5B and 6B) due to higher amounts of precipitation and because monsoon representation in the S2S system is not ideal. T2M error decreases in these months (Figure 5A and 6A) because air temperature is most strongly related to ENSO during the monsoon season (Zhou et al., 2019) and GEOS-S2S tends to capture ENSO rather well (Molod et al., 2020; Hackert et al., 2020; Lim et al., 2021). In a regional sense, Figure 3AB show that errors in precipitation are generally higher for the South subregion, possibly due to the difficulties of accurately forecasting monsoon activity, which can impact precipitation in the Indian subcontinent more than other surrounding regions. For snow variables, fSCA and SWE have low errors when snow is low during the warmest months (Figure 5CD and Figure 6C). An exception is shown in Figure 6D, where forecast errors for SWE are higher during the warm months and lower in the fall, which could be due to the forecasts accumulating SWE more rapidly in the S2S system than what is shown in the HMA-SR reanalysis product. Another explanation for this could be the role of westerly disturbances, which bring enhanced precipitation during the winter months for the west and northern parts of HMA (Cannon et al., 2016), where in our analysis the precipitation forecasts for these regions are underconfident (Figure S1) and larger errors for fSCA (Figure 3CD and Figure 9CD) and SWE (Figure 3CD and Figure 10CD) can be expected in these regions. For SM and TWS, error patterns in Figure 5EF and Figure 6EF may be related to monsoon representation in the S2S system,

but the errors can also be associated with the observational difference in the seasonal cycles shown in Figures 4E and 4F. Overall, improving the representation of monsoon and westerly dynamics in GEOS-S2S may improve forecast skill, particularly during and following the monsoon season.

Our results confirm those from recent studies, such as Deoras et al. (2021), who compared the predictions of the Indian Monsoon low pressure system in various S2S prediction models on a time scale of 15 days to ERA-Interim and MERRA-2 reanalysis data. Their study found that most models were able to predict basic features, however all S2S models underestimate the frequency of the low-pressure systems and that precipitation biases increased with forecast lead time. Hsu et al. (2021) simulated the East Asian winter monsoon on S2S timescales for 45-day hindcasts using the Model for Prediction Across Scales (MPAS). Their evaluation results revealed that MPAS can simulate the climatological characteristics of the monsoon reasonably, with a surface cold bias for temperature and a positive rainfall bias over East Asia. However, they also found that a biased sea surface temperature may modify the circulation over the Western Pacific and affect the simulated occurrence frequency of cold events near Taiwan during winter. Furthermore, climate models are notoriously known to simulate a double Intertropical convergence Zone (ITCZ), in which excessive precipitation is produced on both sides of the equator and especially in the Southern Hemisphere tropics (Hwang and Dargan 2013; Zhang et al., 2019). This is a problem that has been persistent in climate model simulations and can impact the results of S2S forecasts in the HMA region.

4.3.3 Representation of Land Processes

Differences in the level of the S2S forecast skill relative to MERRA-2 and to the other reference products (Table 2 and Figure 2) could be due to certain physical processes that are seen in the signatures of the reference data products but under-represented in the frameworks of GEOS-S2S and MERRA-2. Characterizing hydrometeorological conditions in HMA, through both observations and modeling, is difficult owing to the scarcity of in-situ observations and the complex orographic conditions that impede accurate retrievals of satellite estimates and due to properly representing these processes in the model simulations (Su et al., 2013; Ghatak et al., 2018; Loomis et al., 2019a; Yoon et al., 2019; Gerlitz et al., 2020). These challenges are reflected in the wide range of GEOS-S2S forecast skill when compared to MERRA-2 and reference datasets (i.e., as seen in Figures 2, 4-12).

For example, ESA-CCI data of SM are probably of limited quality in the topographically complex HMA region, and GRACE TWS data shows the signature from rivers, lakes, glacier mass changes, and groundwater pumping that are included in GRACE data but not fully represented in the GEOS-S2S modeling framework. Some regions within HMA, particularly the Indian subcontinent, are known for intense over-pumping of groundwater, which has led to extreme levels of groundwater depletion and has played a prominent role in the loss of freshwater storage for these regions (Tiwari et al., 2009; Xiang et al., 2016; Girotto et al., 2017). This dynamic is captured in the GRACE data but not in the GEOS-S2S

forecasts (i.e., compare Figure 7C and Figure 7D). More realistic representation of the various water budget components within GEOS-S2S, such as surface water or groundwater pumping, is likely to contribute to improved skill in the S2S forecasts.

Appropriate representation of seasonal snow along with temperatures and antecedent precipitation are critical to realistically forecast the HMA energy and water cycles. GEOS-S2S forecasts tend to underestimate temperature and overestimate precipitation relative to both MERRA-2 and the reference observations during all months and nearly all lead times (Figure 4A-B); this cumulatively impacts snow cover and volume (Figure 4C-D). MERRA-2 corrected precipitation has a known dry bias (Figure 4B; Yoon et al., 2019), which limits fSCA and SWE accumulation in the MERRA-2 product (Figure 4C-D). GEOS-S2S is initialized with similar land conditions to MERRA-2, resulting in low fSCA and SWE during winter 1-month lead forecasts; however, GEOS-S2S atmospheric physics increases precipitation as forecasts continue for 2- and 3-month lead forecasts, with more extensive snow cover and higher snow volume (Figure 4C-D), resulting in a seasonal cycle that more closely approximates MODIS and the HMA-SR. This results in a relatively constant regional ubRMSE for all lead times when compared to MODIS and the HMA-SR (Figure 6C-D) and localized improvements in ubRMSE with lead time across the Hindu Kush and Karakoram (Figure 9H-L and Figure 10H-L).

Despite the improvement in the absolute magnitude of snow volume due to increasing precipitation, limitations in the snow depletion curve used within GEOS-S2S and MERRA-2 result in more extensive snow coverage regionally and more limited reduction in fSCA relative to SWE in the Hindu Kush (Figure 9H-L and Figure 10H-L). Both GEOS-S2S and MERRA-2 systems use a globally consistent linear relationship between SWE and fSCA with the minimum SWE needed to fully cover a pixel in snow; that is fSCA=1 if SWE is greater than 26 mm (Stieglitz et al., 2001; Toure et al., 2018). This prescription was developed based on studies in the northeastern USA and oversimplifies the relationship between SWE and fSCA in mountainous regions (e.g., Schneider et al., 2021) and results in too much snow cover in the GEOS-S2S forecasts (Figure 4C). Considering the regional pattern of the SWE-fSCA relationship, in addition to improvements in topography (section 4.2.1) and inclusion of regionally important processes like surface albedo evolution, through assimilation (Girotto et al., 2020) or directly modeling aerosol deposition on snow (Sarangi et al., 2019, 2020), will likely improve snow forecasting and associated runoff from snow melt within GEOS-S2S

4.4 S2S Forecasting for Society's Needs

There are various efforts in the broader community (e.g., Arendt et al., 2017), that are aimed at addressing climate change impacts on natural hazards (such as flooding or landslides) in the HMA region. S2S predictions for HMA from GEOS-S2S can potentially provide useful information for the local populations, for example by potentially providing forecasts with several months lead time that can be beneficial in preparing for local natural hazards (Bekaert et al., 2020; Stanley et al., 2020). Different subregions within HMA can benefit in different ways from S2S forecasts, based on the varying needs of

local populations. Studies that utilize numerical methods and state-of-the-art model initialization to enhance S2S prediction skill are beginning to emerge. For example, Gerlitz et al. (2020) provided a review of seasonal forecasts of water availability in Central Asia. Their review showed that exceptionally skillful discharge forecasts for the agriculturally relevant vegetation season can be derived by means of statistical models taking remote sensing-based estimations of snow coverage in the Central Asian mountain regions as independent covariates, and they found that the consideration of global climate indices, in particular El Niño, allows to extend the forecast lead-times. Therefore, there is reason to believe that improvements in S2S forecast skill can generally be achieved.

In our study, the modest levels of forecast error provide a sense of trust in the model forecasts in the context of S2S forecasting skill. For example, when compared to MERRA-2, the anomaly correlation for forecasts at 1-month lead was above 0.18 for all variables and as high as 0.62 (for SM). Relative to the reference data products, the anomaly correlation for forecasts at 1-month lead was above 0.13 for all variables and as high as 0.24 (for fSCA). Compared to other S2S evaluation studies, these results for HMA are promising. For instance, de Andrade et al., (2019) showed that anomaly correlation of global precipitation forecasts at lead time of 1 to 4 weeks was greatly reduced with lead time for a variety of S2S models, and by week 4 the anomaly correlation was consistently below 0.2 for all models. This skill level is comparable to the results presented here for the GEOS-S2S forecasts in the HMA region. Therefore, the GEOS-S2S forecasts for HMA shown in our study generally have acceptable skill at 1-month lead time compared to other S2S studies. For this study, we use multiple sources of observed and verification data to estimate the forecast skill since relying on solely one source of information may be misleading. Here, we used two different products for each climate variable to get a sense of the uncertainty in the forecast skill for each variable. We are not aware of a merged data product for the HMA region, which would be extremely valuable for an evaluation study like this, but perhaps the combination of MERRA-2 data with other verification data products is a good alternative.

Results shown here, and from the GEOS-S2S system in general, can help the community benchmark the S2S forecasting skill for the HMA region, and for specific subregions within HMA, and can also help the community synthesize areas of model improvements that can potentially enhance the forecast skill or expand the time horizon of skillful forecasts. Other areas of enhancing the S2S forecasts could be achieved by the assimilation of land surface observations during the initialization period, for variables such as surface soil moisture (Koster et al., 2011) or snow-covered area (Senan et al., 2016). More accurate representation of initial conditions could lead to improved forecast accuracy at the 1-month lead time, but it is possible that a gain in skill can persist for 2- and 3-month lead times, and perhaps longer. Given the confluence of water resource needs from the local population and the complexity of the hydrologic cycle in HMA, further investment for improving S2S forecasts can be extremely useful for this region, and such improvements can potentially be felt globally.

5 Conclusions

We showed here an evaluation of the GEOS-S2S forecasting system in the HMA region, utilizing various products such as reanalysis data as well as data sets obtained from satellites or model data fusion products. The hydrometeorological variables in our evaluation results included 2-m air temperature (T2M), total precipitation (PRECTOT), fractional snow cover area (fSCA), snow water equivalent (SWE), surface soil moisture (SM), and terrestrial water storage (TWS). The main data product used for the evaluation was the MERRA-2 reanalysis product, which provided information to compare all the considered variables in GEOS-S2S. For further verification, we used separate data for the evaluation of each variable, including ERA5 for T2M, APHRODITE for PRECTOT, MODIS for fSCA, HMA-SR for SWE, ESA-CCI for SM, and GRACE for TWS. We showed various aspects of the model evaluation, such as the skill based on variables, lead time, or observation used for the evaluation. To gain a more regional point of view, we showed the evaluation based on specific subregions. We also displayed the climatology of the GEOS-S2S ensemble mean, the ensemble spread, and the mean error for each variable.

Choice of evaluation datasets heavily impacted our results. For example, when compared to MERRA-2, variables with longer memory in the physical climate system, such as soil moisture and TWS, had higher accuracy in the S2S forecasting system compared to variables representing quickly changing processes, such as temperature or precipitation. This was true when comparing the S2S forecasts to the MERRA-2 reanalysis because of similar initialization and model architecture as used in GEOS-S2S. However, this finding was not conclusive when reference data products were used in the evaluation. Finally, we provided potential avenues for model improvements that can help enhance the forecasts, such as higher resolution topography representation as well as more realistic representation of surface water and groundwater pumping. These improvements can help, for example, with forecasts of TWS since the model does not have groundwater pumping whereas the GRACE signature includes this process. Other paths to improvement could be the assimilation of observations for the initialization of land surface state variables, such as soil moisture or snow cover.

Our results shown here benchmark the GEOS-S2S system's ability to forecast HMA on the 1-3 month timescale. We showed that, when compared to MERRA-2, the anomaly correlation for forecasts at 1-month lead was above 0.18 for all variables and as high as 0.62 (for SM). Relative to the reference data products, the anomaly correlation for forecasts at 1-month lead was above 0.13 for all variables and as high as 0.24 (for fSCA). Compared to other S2S evaluation studies, these results for HMA are promising. The reported results should motivate future improvements in the forecasts, such as model initialization, model physics, or more realistic orographic representation, that will be helpful for climate adaptation, natural hazard mitigation, and water resources planning for the population of HMA.

6 Code Availability

MATLAB code and other scripts to process the GEOS-S2S outputs are available on request, and scripts to plot and evaluate the data and produce all analysis in this study are available upon request from Dr. Massoud.

7 Data Availability

The GEOS-S2S-V2 data is available on the Discover server of NCCS. GEOS-S2S-V2 forecast output data are presently available at https://gmao.gsfc.nasa.gov/gmaoftp/gmaofcst/. The file specification document that elaborates on the available output from GEOS-S2S is available online (from https://gmao.gsfc.nasa.gov/pubs/docs/Nakada1033.pdf). MERRA-2 data can be downloaded at no cost from (https://disc.gsfc.nasa.gov/datasets?project=MERRA-2), ERA5 data from (https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels-monthly-means?tab=overview), MODIS data from (https://modis.gsfc.nasa.gov/data/), HMA-SR data from (https://nsidc.org/data/HMA_SR_D/versions/1), ESA-CCI from (https://esa-soilmoisture-cci.org/), and GRACE data from (https://earth.gsfc.nasa.gov/geo/data/grace-mascons).

8 Author Contribution

Dr. Massoud, Dr. Andrews, Dr. Girotto – Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Writing – original draft preparation, Writing – review and editing, Visualization, Validation. Dr. Kim and S. Ruehr – Data Curation, Writing – review and editing. Dr. Reichle and Dr. Molod – Formal Analysis, Investigation, Methodology, Supervision, Writing – review and editing.

9 competing Interest

The authors declare that they have no conflict of interest.

10 Acknowledgements

The authors acknowledge the NASA HiMAT team for funding this work (GRANT # 80NSSC20K1301), as well as for generous data sharing and broader discussions that helped shape the paper. GMAO's GEOS-S2S-V2 development was funded under the NASA Modeling, Analysis, and Prediction program GMAO “core” funding. Computational resources were provided by the NASA High-End Computing (HEC) Program through the NASA Center for Climate Simulation (NCCS) at the Goddard Space Flight Center and the NASA Advanced Supercomputing (NAS) division. The authors thank Agniv Sengupta for information on and preparation of the topography map. This manuscript has been authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The US government retains and the

publisher, by accepting the article for publication, acknowledges that the US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan.

References

Arendt, Anthony A., Paul Houser, Sarah B. Kapnick, Jeffrey S. Kargel, Dalia Kirschbaum, Sujay Kumar, Steven A. Margulis et al. "NASA's High Mountain Asia Team (HiMAT): collaborative research to study changes of the High Asia region." In AGU Fall Meeting Abstracts, vol. 2017, pp. C33D-1231. Bibcode: 2017AGUFM.C33D1231A, 2017.

Aquila, Valentina, Colleen Baldwin, Nikita Mukherjee, Eric Hackert, Feng Li, Jelena Marshak, Andrea Molod, and Steven Pawson. "Impacts of the Eruption of Mount Pinatubo on Surface Temperatures and Precipitation Forecasts With the NASA GEOS Subseasonal-to-Seasonal System." Journal of Geophysical Research: Atmospheres 126, no. 16. e2021JD034830. doi.org/10.1029/2021JD034830, 2021.

Batbaatar, Jigjidsurengiin, Alan R. Gillespie, Michele Koppes, Douglas H. Clark, Oliver A. Chadwick, David Fink, Ari Matmon, and Summer Rupper. "Glacier development in continental climate regions of central Asia." Untangling the Quaternary Period: A Legacy of Stephen C. Porter 548: 123. doi.org/10.1130/2020.2548(07), 2021.

Bekaert, David PS, Alexander L. Handwerger, Piyush Agram, and Dalia B. Kirschbaum. "InSAR-based detection method for mapping and monitoring slow-moving landslides in remote regions with steep and mountainous terrain: An application to Nepal." Remote Sensing of Environment 249: 111983. doi.org/10.1016/j.rse.2020.111983, 2020.

Benjamin, Stanley G., John M. Brown, Gilbert Brunet, Peter Lynch, Kazuo Saito, and Thomas W. Schlatter. "100 years of progress in forecasting and NWP applications." Meteorological Monographs 59: 13-1. doi.org/10.1175/AMSMONOGRAPHS-D-18-0020.1, 2019.

Bosilovich, M. G., R. Lucchesi, and M. Suarez: MERRA-2: File Specification. GMAO Office Note No. 9 (Version 1.1), 73 pp, available from http://gmao.gsfc.nasa.gov/pubs/office_notes, 2016.

Cannon, F., Carvalho, L., Jones, C. and Norris, J., 2016. Winter westerly disturbance dynamics and precipitation in the western Himalaya and Karakoram: a wave-tracking approach. Theoretical and Applied Climatology, 125(1), pp.27-44. DOI 10.1007/s00704-015-1489-8.

Cannon, Forest, Leila MV Carvalho, Charles Jones, Jesse Norris, Bodo Bookhagen, and George N. Kiladis. "Effects of topographic smoothing on the simulation of winter precipitation in High Mountain Asia." Journal of Geophysical Research: Atmospheres 122, no. 3: 1456-1474. doi.org/10.1002/2016JD026038, 2017.

Cavalieri,D. J., Parkinson, C. L., Gloersen, P.,&Zwally,H. J. (1996). Sea ice concentrations from Nimbus-7SMMRandDMSP SSM/I-SSMIS Passive Microwave Data, Version 1. [1978–2017]. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center, https://doi.org/10.5067/8GQ8LZQVL0VL.

Christensen, Michael F., Matthew J. Heaton, Summer Rupper, C. Shane Reese, and William F. Christensen. "Bayesian Multi-Scale Spatio-Temporal Modeling of Precipitation in the Indus Watershed." Frontiers in Earth Science 7: 210. doi.org/10.3389/feart.2019.00210, 2019.

Dars, Ghulam Hussain, Courtenay Strong, Adam K. Kochanski, Kamran Ansari, and Syed Hammad Ali. "The spatiotemporal variability of temperature and precipitation over the upper Indus Basin: An evaluation of 15 year WRF simulations." Applied Sciences 10, no. 5: 1765. doi.org/10.3390/app10051765, 2020.

de Andrade, Felipe M., Caio AS Coelho, and Iracema FA Cavalcanti. "Global precipitation hindcast quality assessment of the Subseasonal to Seasonal (S2S) prediction project models." Climate Dynamics 52, no. 9: 5451-5475. doi.org/10.1007/s00382-018-4457-z, 2019.

De Lannoy, Gabriëlle JM, and Rolf H. Reichle. "Global assimilation of multiangle and multipolarization SMOS brightness temperature observations into the GEOS-5 catchment land surface model for soil moisture estimation." Journal of Hydrometeorology 17, no. 2: 669-691. doi.org/10.1175/JHM-D-15-0037.1, 2016.

de Rosnay, Patricia, Gianpaolo Balsamo, Clément Albergel, Joaquín Muñoz-Sabater, and Lars Isaksen. "Initialisation of land surface variables for numerical weather prediction." Surveys in Geophysics 35, no. 3: 607-621. doi.org/10.1007/s10712-012-9207-x, 2014.

DeFlorio, Michael J., Duane E. Waliser, Bin Guan, F. Martin Ralph, and Frédéric Vitart. "Global evaluation of atmospheric river subseasonal prediction skill." Climate Dynamics 52, no. 5: 3039-3060. doi.org/10.1007/s00382-018-4309-x, 2019.

Deoras, A., Hunt, K. M. R., & Turner, A. G. Comparison of the Prediction of Indian Monsoon Low Pressure Systems by Subseasonal-to-Seasonal Prediction Models, Weather and Forecasting, 36(3), 859-877. doi.org/10.1175/WAF-D-20-0081.1, 2021.

Ding, Qinghua, and Bin Wang. "Intraseasonal teleconnection between the summer Eurasian wave train and the Indian monsoon." Journal of Climate 20, no. 15: 3751-3767. doi.org/10.1175/JCLI4221.1, 2007.

Dirmeyer, Paul A., Subhadeep Halder, and Rodrigo Bombardi. "On the harvest of predictability from land states in a global forecast model." Journal of Geophysical Research: Atmospheres 123, no. 23: 13-111. doi.org/10.1029/2018JD029103, 2018.

Dorigo, Wouter, Wolfgang Wagner, Clement Albergel, Franziska Albrecht, Gianpaolo Balsamo, Luca Brocca, Daniel Chung et al. "ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions." Remote Sensing of Environment 203: 185-215. doi.org/10.1016/j.rse.2017.07.001, 2017.

Famiglietti, J.S.; Lo, M.-H.; Ho, S.L.; Bethune, J.; Anderson, K.J.; Syed, T.H.; Swenson, S.C.; De Linage, C.R.; Rodell, M. Satellites measure recent rates of groundwater depletion in California's Central Valley. Geophys. Res. Lett., 38, 471. doi.org/10.1029/2010GL046442, 2011.

Fortin, V., Abaza, M., Anctil, F. and Turcotte, R., 2014. Why should ensemble spread match the RMSE of the ensemble mean?. Journal of Hydrometeorology, 15(4), pp.1708-1713. DOI: 10.1175/JHM-D-14-0008.1.

Gelaro, Ronald, Will McCarty, Max J. Suárez, Ricardo Todling, Andrea Molod, Lawrence Takacs, Cynthia A. Randles et al. "The modern-era retrospective analysis for research and applications, version 2 (MERRA-2)." Journal of climate 30, no. 14: 5419-5454. doi.org/10.1175/JCLI-D-16-0758.1, 2017.

Gerlitz, Lars, Sergiy Vorogushyn, and Abror Gafurov. "Climate informed seasonal forecast of water availability in Central Asia: State-of-the-art and decision making context." Water Security 10: 100061. doi.org/10.1016/j.wasec.2020.100061, 2020.

Getirana, Augusto, Hahn Chul Jung, Kristi Arsenault, Shraddhanand Shukla, Sujay Kumar, Christa Peters-Lidard, Issoufou Maigari, and Bako Mamane. "Satellite gravimetry improves seasonal streamflow forecast initialization in Africa." Water Resources Research 56, no. 2: e2019WR026259. doi.org/10.1029/2019WR026259, 2020.

Ghatak, Debjani, Benjamin Zaitchik, Sujay Kumar, Mir A. Matin, Birendra Bajracharya, Christopher Hain, and Martha Anderson. "Influence of precipitation forcing uncertainty on hydrological simulations with the NASA South Asia land data assimilation system." Hydrology 5, no. 4: 57. doi.org/10.3390/hydrology5040057, 2018.

Gibson, Peter B., Duane E. Waliser, Alexander Goodman, Michael J. DeFlorio, Luca Delle Monache, and Andrea Molod. "Subseasonal-to-Seasonal Hindcast Skill Assessment of Ridging Events Related to Drought Over the Western United States." Journal of Geophysical Research: Atmospheres 125, no. 22: e2020JD033655. doi.org/10.1029/2020JD033655, 2020.

Girotto, M., Margulis, S.A. and Durand, M., 2014. Probabilistic SWE reanalysis as a generalization of deterministic SWE reconstruction techniques. Hydrological processes, 28(12), pp.3875-3895. https://doi.org/10.1002/hyp.9887.

Girotto, Manuela, Gabriëlle JM De Lannoy, Rolf H. Reichle, Matthew Rodell, Clara Draper, Soumendra N. Bhanja, and Abhijit Mukherjee. "Benefits and pitfalls of GRACE data assimilation: A case study of terrestrial water storage depletion in India." Geophysical research letters 44, no. 9: 4107-4115. doi.org/10.1002/2017GL072994, 2017.

Girotto, Manuela, Keith N. Musselman, and Richard LH Essery. "Data assimilation improves estimates of climate-sensitive seasonal snow." Current Climate Change Reports 6, no. 3: 81-94. doi.org/10.1007/s40641-020-00159-7, 2020.

Global Modeling and Assimilation Office (GMAO), MERRA-2 tavgM_2d_slv_Nx: 2d, Monthly mean, Time-Averaged, Single-Level, Assimilation, Single-Level Diagnostics V5.12.4, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed: [August 2021], 10.5067/AP1B0BA5PD2K, 2015a.

Global Modeling and Assimilation Office (GMAO), MERRA-2 tavgM_2d_flx_Nx: 2d, Monthly mean, Time-Averaged, Single-Level,Assimilation, Surface Flux Diagnostics V5.12.4, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed: [August 2021], 10.5067/0JRLVL8YV2Y4, 2015b.

Global Modeling and Assimilation Office (GMAO), MERRA-2 tavgM_2d_lnd_Nx: 2d, Monthly mean, Time-Averaged, Single-Level, Assimilation, Land Surface Diagnostics V5.12.4, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed: [August 2021], 10.5067/8S35XF81C28F, 2015c.

Griffies, Stephen M. "Elements of the modular ocean model (MOM)." GFDL Ocean Group Tech. Rep 7, no. 620 (2012): 47.

Gruber, A., Scanlon, T., van der Schalie, R., Wagner, W., and Dorigo, W.: 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, 2019.

Hatsuzuka, Daisuke, and Hatsuki Fujinami. "Effects of the South Asian monsoon intraseasonal modes on genesis of low pressure systems over Bangladesh." Journal of Climate 30, no. 7: 2481-2499. doi.org/10.1175/JCLI-D-16-0360.1, 2017.

Hackert, Eric, Robin M. Kovach, A. Molod, G. Vernieres, A. Borovikov, J. Marshak, and Y. Chang. "Satellite sea surface salinity observations impact on El Niño/Southern Oscillation predictions: Case studies from the NASA GEOS seasonal forecast system." Journal of Geophysical Research: Oceans 125, no. 4: e2019JC015788. doi.org/10.1029/2019JC015788, 2020.

Hall, Dorothy K., George A. Riggs, Vincent V. Salomonson, Nicolo E. DiGirolamo, and Klaus J. Bayr. "MODIS snow-cover products." Remote sensing of Environment 83, no. 1-2: 181-194. doi.org/10.1016/S0034-4257(02)00095-0, 2002.

Hall, D. K. and G. A. Riggs. MODIS/Terra Snow Cover Daily L3 Global 0.05Deg CMG, Version 6. [MOD10A1]. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. doi: https://doi.org/10.5067/MODIS/MOD10C1.006. Accessed: [December 2020]. 2016a.

Hall, D. K. and G. A. Riggs. MODIS/Terra Snow Cover Daily L3 Global 500m SIN Grid, Version 6. [MOD10A1]. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. doi: https://doi.org/10.5067/MODIS/MOD10A1.006. Accessed: [December 2020]. 2016b.

Hill C, DeLuca C, Suarez M, Da Silva AR. The architecture of the earth system modeling framework. Computing in Science & Engineering. Aug 2;6(1):18-28. 10.1109/MCISE.2004.1255817, 2004.

Hersbach, Hans, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas et al. "The ERA5 global reanalysis." Quarterly Journal of the Royal Meteorological Society 146, no. 730: 1999-2049. doi.org/10.1002/qj.3803, 2020.

Hopson, T.M., 2014. Assessing the ensemble spread–error relationship. Monthly Weather Review, 142(3), pp.1125-1142. https://doi.org/10.1175/MWR-D-12-00111.1.

Hsu, L.-H.; Chen, D.-R.; Chiang, C.-C.; Chu, J.-L.; Yu, Y.-C.; Wu, C.-C. Simulations of the East Asian Winter Monsoon on Subseasonal to Seasonal Time Scales Using the Model for Prediction Across Scales. Atmosphere, 12, 865. doi.org/10.3390/atmos12070865, 2021.

Hunke, E. C., and W. H. Lipscomb. "The Los Alamos sea ice model documentation and software user's manual, Version 4.0." Los Alamos National Laboratory (2008).

Hwang, Yen-Ting, and Dargan MW Frierson. "Link between the double-Intertropical Convergence Zone problem and cloud biases over the Southern Ocean." Proceedings of the National Academy of Sciences 110, no. 13: 4935-4940. doi.org/10.1073/pnas.1213302110, 2013.

Immerzeel, W.W., Van Beek, L.P. and Bierkens, M.F., 2010. Climate change will affect the Asian water towers. science, 328(5984), pp.1382-1385. DOI: 10.1126/science.1183188

Immerzeel, W.W., Lutz, A.F., Andrade, M. et al. Importance and vulnerability of the world's water towers. Nature 577, 364–369. doi.org/10.1038/s41586-019-1822-y, 2020.

Jiang, Xianan, Tim Li, and Bin Wang. "Structures and mechanisms of the northward propagating boreal summer intraseasonal oscillation." Journal of Climate 17, no. 5: 1022-1039. doi.org/10.1175/1520-0442(2004)017<1022:SAMOTN>2.0.CO;2, 2004.

Kirtman, B.P., Min, D., Infanti, J.M., Kinter, J.L., Paolino, D.A., Zhang, Q., Van Den Dool, H., Saha, S., Mendez, M.P., Becker, E. and Peng, P.. The North American multimodel ensemble: phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. Bulletin of the American Meteorological Society, 95(4), pp.585-601. doi.org/10.1175/BAMS-D-12-00050.1, 2014.

Koster, Randal D., Max J. Suarez, Agnès Ducharne, Marc Stieglitz, and Praveen Kumar. "A catchment-based approach to modeling land surface processes in a general circulation model: 1. Model structure." Journal of Geophysical Research: Atmospheres 105, no. D20: 24809-24822. doi.org/10.1029/2000JD900327, 2000.

Koster, R.D., Mahanama, S.P.P., Yamada, T.J., Balsamo, G., Berg, A.A., Boisserie, M., Dirmeyer, P.A., Doblas-Reyes, F.J., Drewitt, G., Gordon, C.T. and Guo, Z., 2011. The second phase of the global land–atmosphere coupling experiment: soil moisture contributions to subseasonal forecast skill. Journal of Hydrometeorology, 12(5), pp.805-822. https://doi.org/10.1175/2011JHM1365.1.

Li, Jian, Rucong Yu, Tianjun Zhou, and Bin Wang. "Why is there an early spring cooling shift downstream of the Tibetan Plateau?." Journal of Climate 18, no. 22: 4660-4668. doi.org/10.1175/JCLI3568.1, 2005.

Li, Jian, Rucong Yu, and Tianjun Zhou. "Teleconnection between NAO and climate downstream of the Tibetan Plateau." Journal of Climate 21, no. 18: 4680-4690. doi.org/10.1175/2008JCLI2053.1, 2008.

Lim, Young-Kwon. "The East Atlantic/West Russia (EA/WR) teleconnection in the North Atlantic: climate impact and relation to Rossby wave propagation." Climate dynamics 44, no. 11: 3211-3222. doi.org/10.1007/s00382-014-2381-4, 2015.

Lim, Young-Kwon, Nathan P. Arnold, Andrea M. Molod, and Steven Pawson. "Seasonality in Prediction Skill of the Madden-Julian Oscillation and Associated Dynamics in Version 2 of NASA's GEOS-S2S Forecast System." Journal of Geophysical Research: Atmospheres 126, no. 18: e2021JD034961. doi.org/10.1029/2021JD034961, 2021.

Liu, Yufei, and Steven A. Margulis. "Deriving Bias and uncertainty in MERRA-2 snowfall precipitation over High Mountain Asia." Frontiers in Earth Science 7: 280. doi.org/10.3389/feart.2019.00280, 2019.

Liu, Y., Fang, Y., and Margulis, S. A.: Spatiotemporal distribution of seasonal snow water equivalent in High Mountain Asia from an 18-year Landsat–MODIS era snow reanalysis dataset, The Cryosphere, 15, 5261–5280, https://doi.org/10.5194/tc-15-5261-2021, 2021a.

Liu, Y., Y. Fang, and S. A. Margulis. High Mountain Asia UCLA Daily Snow Reanalysis, Version 1. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. doi: https://doi.org/10.5067/HNAUGJQXSCVU. Accessed: [December 2021]. 2021b.

Loomis, Bryant D., Alexandra S. Richey, Anthony A. Arendt, Ravi Appana, Y-JC Deweese, Bart A. Forman, Sujay V. Kumar, Terence J. Sabaka, and David E. Shean. "Water storage trends in high mountain Asia." Frontiers in earth science 7: 235. doi.org/10.3389/feart.2019.00235, 2019a.

Loomis, B.D., Luthcke, S.B. & Sabaka, T.J. Regularization and error characterization of GRACE mascons. J Geod 93, 1381–1398. https://doi.org/10.1007/s00190-019-01252-y, 2019b.

Maeda, Mio, Yasutomi, Natsuko, Yatagai, Akiyo & National Center for Atmospheric Research Staff (Eds). Last modified 29 Jul 2020. "The Climate Data Guide: APHRODITE: Asian Precipitation - Highly-Resolved Observational Data Integration Towards Evaluation of Water Resources." Retrieved from https://climatedataguide.ucar.edu/climate-data/aphrodite-asian-precipitation-highly-resolved-observational-data-integration-towards.

Margulis, Steven A., Yufei Liu, and Elisabeth Baldo. "A joint landsat-and MODIS-based reanalysis approach for midlatitude montane seasonal snow characterization." Frontiers in Earth Science 7: 272. doi.org/10.3389/feart.2019.00272, 2019.

Mariotti, Annarita, Paolo M. Ruti, and Michel Rixen. "Progress in subseasonal to seasonal prediction through a joint weather and climate community effort." npj Climate and Atmospheric Science 1, no. 1: 1-4. doi.org/10.1038/s41612-018-0014-z, 2018.

Mariotti, Annarita, Cory Baggett, Elizabeth A. Barnes, Emily Becker, Amy Butler, Dan C. Collins, Paul A. Dirmeyer et al. "Windows of opportunity for skillful forecasts subseasonal to seasonal and beyond." Bulletin of the American Meteorological Society 101, no. 5: E608-E625. doi.org/10.1175/BAMS-D-18-0326.1, 2020.

Massoud, E.C.; Purdy, A.J.; Miro, M.E.; Famiglietti, J.S. Projecting groundwater storage changes in California's Central Valley. Sci. Rep., 8, 12917. doi.org/10.1038/s41598-018-31210-1, 2018.

Massoud, Elias C., Zhen Liu, Amin Shaban, and Mhamad El Hage. "Groundwater Depletion Signals in the Beqaa Plain, Lebanon: Evidence from GRACE and Sentinel-1 Data." Remote Sensing 13, no. 5: 915. doi.org/10.3390/rs13050915, 2021.

Massoud, E. C., Bloom, A. A., Longo, M., Reager, J. T., Levine, P. A., and Worden, J. R.: Information content of soil hydrology in a west Amazon watershed as informed by GRACE, Hydrol. Earth Syst. Sci., 26, 1407–1423, https://doi.org/10.5194/hess-26-1407-2022, 2022.

Maurer, J. M., J. M. Schaefer, J. B. Russell, S. Rupper, N. Wangdi, A. E. Putnam, and N. Young. "Seismic observations, numerical modeling, and geomorphic analysis of a glacier lake outburst flood in the Himalayas." Science advances 6, no. 38: eaba3645. DOI: 10.1126/sciadv.aba3645, 2020.

Meena, Narendra Kumar, Pranaya Diwate, and Sundeep Pandita. "Evidence of ENSO and IOD Interplay in Continental Climatic Records from Southern Himalaya (Renuka Lake), India." (2022).

Merryfield, W. J., Baehr, J., Batté, L., Becker, E. J., Butler, A. H., Coelho, C. A. S., Danabasoglu, G., Dirmeyer, P. A., Doblas-Reyes, F. J., Domeisen, D. I. V., Ferranti, L., Ilynia, T., Kumar, A., Müller, W. A., Rixen, M., Robertson, A. W., Smith, D. M., Takaya, Y., Tuma, M., Vitart, F., White, C. J., Alvarez, M. S., Ardilouze, C., Attard, H., Baggett, C., Balmaseda, M. A., Beraki, A. F., Bhattacharjee, P. S., Bilbao, R., de Andrade, F. M., DeFlorio, M. J., Díaz, L. B., Ehsan, M. A., Fragkoulidis, G., Grainger, S., Green, B. W., Hell, M. C., Infanti, J. M., Isensee, K., Kataoka, T., Kirtman, B. P., Klingaman, N. P., Lee, J., Mayer, K., McKay, R., Mecking, J. V., Miller, D. E., Neddermann, N., Justin Ng, C. H., Ossó, A., Pankatz, K., Peatman, S., Pegion, K., Perlwitz, J., Recalde-Coronel, G. C., Reintges, A., Renkl, C., Solaraju-Murali, B., Spring, A., Stan, C., Sun, Y. Q., Tozer, C. R., Vigaud, N., Woolnough, S., & Yeager, S. Current and Emerging

Developments in Subseasonal to Decadal Prediction, Bulletin of the American Meteorological Society, 101(6), E869-E896. doi.org/10.1175/BAMS-D-19-0037.1, 2020.

Mishra, Shruti K., Thomas D. Veselka, Alexander A. Prusevich, Danielle S. Grogan, Richard B. Lammers, David R. Rounce, Syed H. Ali, and Mark H. Christian. "Differential impact of climate change on the hydropower economics of two river basins in high mountain Asia." Frontiers in Environmental Science 8: 26. doi.org/10.3389/fenvs.2020.00026, 2020.

Molod, A., Takacs, L., Suarez, M., and Bacmeister, J.: Development of the GEOS-5 atmospheric general circulation model: evolution from MERRA to MERRA2, Geosci. Model Dev., 8, 1339–1356, https://doi.org/10.5194/gmd-8-1339-2015, 2015.

Molod, Andrea, Eric Hackert, Yury Vikhliaev, Bin Zhao, Donifan Barahona, Guillaume Vernieres, Anna Borovikov et al. "GEOS-S2S version 2: The GMAO high-resolution coupled model and assimilation system for seasonal prediction." Journal of Geophysical Research: Atmospheres 125, no. 5. doi.org/10.1029/2019JD031767, 2020.

Nakada, K., R. Kovach, J. Marshak, and A. Molod: S2S-2_1: File Specification. GMAO Office Note No. 16 (Version 1.0), 78 pp, available from http://gmao.gsfc.nasa.gov/pubs/office_notes. 2018.

Nash, Deanna, Leila Carvalho, Charles Jones, and Qinghua Ding. "Winter and spring atmospheric rivers in High Mountain Asia: climatology, dynamics, and variability." Climate Dynamics: 1-23. doi.org/10.1007/s00382-021-06008-z, 2021.

National Academies of Sciences, Engineering, and Medicine. Next Generation Earth System Prediction: Strategies for Subseasonal to Seasonal Forecasts. Washington, DC: The National Academies Press. https://doi.org/10.17226/21873, 2016.

Parajka, J., and G. Blöschl. "The value of MODIS snow cover data in validating and calibrating conceptual hydrologic models." Journal of hydrology 358, no. 3-4: 240-258. doi.org/10.1016/j.jhydrol.2008.06.006, 2008.

Penny, S. G., Kalnay, E., Carton, J. A., Hunt, B. R., Ide, K., Miyoshi, T., and Chepurin, G. A.: The local ensemble transform Kalman filter and the running-in-place algorithm applied to a global ocean general circulation model, Nonlin. Processes Geophys., 20, 1031–1046, https://doi.org/10.5194/npg-20-1031-2013, 2013.

Pielke Sr, Roger A., Glen E. Liston, Joseph L. Eastman, Lixin Lu, and Michael Coughenour. "Seasonal weather prediction as an initial value problem." Journal of Geophysical Research: Atmospheres 104, no. D16: 19463-19479. doi.org/10.1029/1999JD900231, 1999.

Power, Katherine, Josefine Axelsson, Norbu Wangdi, and Qiong Zhang. "Regional and Local Impacts of the ENSO and IOD Events of 2015 and 2016 on the Indian Summer Monsoon—A Bhutan Case Study." Atmosphere 12, no. 8: 954. doi.org/10.3390/atmos12080954, 2021.

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. IEEE Transactions on Geoscience and Remote Sensing, 59(4), pp.2845-2862. doi:10.1109/TGRS.2020.3012896, 2020.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., and Flannery, B.P. Numerical Recipes in C, 2nd Ed., Cambridge University Press, 1992.

Reichle, Rolf H., Q. Liu, Randal D. Koster, Clara S. Draper, Sarith PP Mahanama, and Gary S. Partyka. "Land surface precipitation in MERRA-2." Journal of Climate 30, no. 5: 1643-1664. doi.org/10.1175/JCLI-D-16-0570.1, 2017.

Rienecker, M. M., Todling, R., Bacmeister, J., Takacs, L., Liu, H. C., et al. The GEOS-5 data assimilation system: Documentation of versions 5.0.1 and 5.1.0, and 5.2.0 (NASA Tech. Rep.): Series on Global Modeling and Data Assimilation NASA/TM-2008-104606. 2008.

Robertson, Andrew W., Frederic Vitart, and Suzana J. Camargo. "Subseasonal to seasonal prediction of weather to climate with application to tropical cyclones." Journal of Geophysical Research: Atmospheres 125, no. 6: e2018JD029375. doi.org/10.1029/2018JD029375, 2020.

Rodell, M.; Velicogna, I.; Famiglietti, J.S. Satellite-based estimates of groundwater depletion in India. Nat. Cell Biol. 460, 999–1002. doi.org/10.1038/nature08238, 2009.

Sang, Yan-Fang, Vijay P. Singh, and Kang Xu. "Evolution of IOD-ENSO relationship at multiple time scales." Theoretical and Applied Climatology 136, no. 3: 1303-1309. doi.org/10.1007/s00704-018-2557-7, 2019.

Sarangi, C., Qian, Y., Rittger, K., Bormann, K. J., Liu, Y., Wang, H., Wan, H., Lin, G., and Painter, T. H.: Impact of light-absorbing particles on snow albedo darkening and associated radiative forcing over high-mountain Asia: high-resolution WRF-Chem modeling and new satellite observations, Atmos. Chem. Phys., 19, 7105–7128, https://doi.org/10.5194/acp-19-7105-2019, 2019.

Sarangi, Chandan, Yun Qian, Karl Rittger, L. Ruby Leung, Duli Chand, Kat J. Bormann, and Thomas H. Painter. "Dust dominates high-altitude snow darkening and melt over high-mountain Asia." Nature Climate Change 10, no. 11: 1045-1051. doi.org/10.1038/s41558-020-00909-3, 2020.

Scaife, A.A., Arribas, A., Blockley, E., Brookshaw, A., Clark, R.T., Dunstone, N., Eade, R., Fereday, D., Folland, C.K., Gordon, M. and Hermanson, L. Skillful long-range prediction of European and North American winters. Geophysical Research Letters, 41(7), pp.2514-2519. doi.org/10.1002/2014GL059637, 2014.

Schneider, Dominik, Noah P. Molotch, Jeffrey S. Deems, and Thomas H. Painter. "Analysis of topographic controls on depletion curves derived from airborne lidar snow depth data." Hydrology Research 52, no. 1: 253-265. doi.org/10.2166/nh.2020.267, 2021.

Senan, R., Orsolini, Y.J., Weisheimer, A., Vitart, F., Balsamo, G., Stockdale, T.N., Dutra, E., Doblas-Reyes, F.J. and Basang, D., 2016. Impact of springtime Himalayan–Tibetan Plateau snowpack on the onset of the Indian summer monsoon in coupled seasonal forecasts. Climate Dynamics, 47(9), pp.2709-2725. https://doi.org/10.1007/s00382-016-2993-y.

Shugar, Dan H., Aaron Burr, Umesh K. Haritashya, Jeffrey S. Kargel, C. Scott Watson, Maureen C. Kennedy, Alexandre R. Bevington, Richard A. Betts, Stephan Harrison, and Katherine Strattman. "Rapid worldwide growth of glacial lakes since 1990." Nature Climate Change 10, no. 10: 939-945. doi.org/10.1038/s41558-020-0855-4, 2020.

Shukla, Ravi P., James L. Kinter, and Chul-Su Shin. "Sub-seasonal prediction of significant wave heights over the Western Pacific and Indian Oceans, part II: The impact of ENSO and MJO." Ocean Modelling 123: 1-15. doi.org/10.1016/j.ocemod.2018.01.002, 2018.

Stanley, Thomas, Dalia B. Kirschbaum, Salvatore Pascale, and Sarah Kapnick. "Extreme Precipitation in the Himalayan Landslide Hotspot." In Satellite precipitation measurement, pp. 1087-1111. Springer, Cham. doi.org/10.1007/978-3-030-35798-6_31, 2020.

Stieglitz, Marc, Agnès Ducharne, Randy Koster, and Max Suarez. "The impact of detailed snow physics on the simulation of snow cover and subsurface thermodynamics at continental scales." Journal of Hydrometeorology 2, no. 3: 228-242. doi.org/10.1175/1525-7541(2001)002<0228:TIODSP>2.0.CO;2, 2001.

Stuecker, Malte F., Axel Timmermann, Fei-Fei Jin, Yoshimitsu Chikamoto, Wenjun Zhang, Andrew T. Wittenberg, Esther Widiasih, and Sen Zhao. "Revisiting ENSO/Indian Ocean dipole phase relationships." Geophysical Research Letters 44, no. 5: 2481-2492. doi:10.1002/2016GL072308, 2017.

Suarez, M., Trayanov, A., Hill, C., Schopf, P., & Vikhliaev, Y. MAPL: A high-level programming paradigm to support more rapid and robust encoding of hierarchical trees of interacting high-performance components. In Proceedings of the 2007 Symposium on Component and Framework Technology in High-Performance and Scientific Computing, pp. 11–20. https://doi.org/10.1145/1297385.1297388, 2007.

Su, Fengge, Xiaolan Duan, Deliang Chen, Zhenchun Hao, and Lan Cuo. "Evaluation of the global climate models in the CMIP5 over the Tibetan Plateau." Journal of climate 26, no. 10: 3187-3208. doi.org/10.1175/JCLI-D-12-00321.1, 2013.

Tapley, Byron D., Srinivas Bettadpur, John C. Ries, Paul F. Thompson, and Michael M. Watkins. "GRACE measurements of mass variability in the Earth system." Science 305, no. 5683: 503-505. DOI: 10.1126/science.1099192, 2004.

Tekeli, A. Emre, Zuhal Akyürek, A. Arda Şorman, Aynur Şensoy, and A. Ünal Şorman. "Using MODIS snow cover maps in modeling snowmelt runoff process in the eastern part of Turkey." Remote Sensing of Environment 97, no. 2: 216-230. doi.org/10.1016/j.rse.2005.03.013, 2005.

Tiwari, V.M.; Wahr, J.; Swenson, S. Dwindling groundwater resources in northern India, from satellite gravity observations. Geophys. Res. Lett. 36, 18401. doi.org/10.1029/2009GL039401, 2009.

Toure, Ally M., Rolf H. Reichle, Barton A. Forman, Augusto Getirana, and Gabrielle JM De Lannoy. "Assimilation of MODIS snow cover fraction observations into the NASA catchment land surface model." Remote sensing 10, no. 2: 316. doi.org/10.3390/rs10020316, 2018.

Vitart, Frédéric, and Andrew W. Robertson. "The sub-seasonal to seasonal prediction project (S2S) and the prediction of extreme events." npj Climate and Atmospheric Science 1, no. 1: 1-7. doi.org/10.1038/s41612-018-0013-0, 2018.

Vitart, Frédéric, and Andrew W. Robertson. "Introduction: Why Sub-seasonal to seasonal prediction (S2S)?" In Sub-Seasonal to Seasonal Prediction, pp. 3-15. Elsevier. doi.org/10.1016/B978-0-12-811714-9.00001-2, 2019.

Viviroli, D., Dürr, H.H., Messerli, B., Meybeck, M. and Weingartner, R., 2007. Mountains of the world, water towers for humanity: Typology, mapping, and global significance. Water resources research, 43(7). https://doi.org/10.1029/2006WR005653

Waliser, D., Weickmann, K., Dole, R., Schubert, S., Alves, O., Jones, C., Newman, M., Pan, H.-L., Roubicek, A., Saha, S., Smith, C., vab deb Dool, H., Vitart, F., Wheeler, M., & Whitaker, J. The Experimental MJO Prediction Project. Bulletin of the American Meteorological Society, 87(4), 425–431. http://www.jstor.org/stable/26217155, 2006.

Waliser, Duane, K. Sperber, H. Hendon, D. Kim, E. Maloney, M. Wheeler, K. Weickmann et al. "MJO simulation diagnostics." Journal of Climate 22, no. 11: 3006-3030. doi.org/10.1175/2008JCLI2731.1, 2009.

Wang, Xuanxuan, Liu Liu, Qiankun Niu, Hao Li, and Zongxue Xu. "Multiple Data Products Reveal Long-Term Variation Characteristics of Terrestrial Water Storage and Its Dominant Factors in Data-Scarce Alpine Regions." Remote Sensing 13, no. 12: 2356. doi.org/10.3390/rs13122356, 2021.

White, C. J., Domeisen, D. I. V., Acharya, N., Adefisan, E. A., Anderson, M. L., Aura, S., Balogun, A. A., Bertram, D., Bluhm, S., Brayshaw, D. J., Browell, J., Büeler, D., Charlton-Perez, A., Chourio, X., Christel, I., Coelho, C. A. S., DeFlorio, M. J., Delle Monache, L., Di Giuseppe, F., García-Solórzano, A. M., Gibson, P. B., Goddard, L., González Romero, C., Graham, R. J., Graham, R. M., Grams, C. M., Halford, A., Katty Huang, W. T., Jensen, K., Kilavi, M., Lawal, K. A., Lee, R. W., MacLeod, D., Manrique-Suñén, A., Martins, E. S. P. R., Maxwell, C. J., Merryfield, W. J., Muñoz, Á. G., Olaniyan, E., Otieno, G., Oyedepo, J. A., Palma, L., Pechlivanidis, I. G., Pons, D., Ralph, F. M., Reis, D. S., Jr., Remenyi, T. A., Risbey, J. S., Robertson, D. J. C., Robertson, A. W., Smith, S., Soret, A., Sun, T., Todd, M. C., Tozer, C. R., Vasconcelos, F. C., Jr., Vigo, I., Waliser, D. E., Wetterhall, F., & Wilson, R. G. Advances in the application and utility of subseasonal-to-seasonal predictions, Bulletin of the American Meteorological Society (published online ahead of print 2021). Retrieved Apr 21, 2022, from https://journals.ametsoc.org/view/journals/bams/aop/BAMS-D-20-0224.1/BAMS-D-20-0224.1.xml. doi.org/10.1175/BAMS-D-20-0224.1, 2021.

Xiang, Longwei, Hansheng Wang, Holger Steffen, Patrick Wu, Lulu Jia, Liming Jiang, and Qiang Shen. "Groundwater storage changes in the Tibetan Plateau and adjacent areas revealed from GRACE satellite gravity data." Earth and Planetary Science Letters 449: 228-239. doi.org/10.1016/j.epsl.2016.06.002, 2016.

Yatagai, Akiyo, Kenji Kamiguchi, Osamu Arakawa, Atsushi Hamada, Natsuko Yasutomi, and Akio Kitoh. "APHRODITE: Constructing a long-term daily gridded precipitation dataset for Asia based on a dense network of rain gauges." Bulletin of the American Meteorological Society 93, no. 9: 1401-1415. doi.org/10.1175/BAMS-D-11-00122.1, 2012.

Yoon, Yeosang, Sujay V. Kumar, Barton A. Forman, Benjamin F. Zaitchik, Yonghwan Kwon, Yun Qian, Summer Rupper et al. "Evaluating the uncertainty of terrestrial water budget components over High Mountain Asia." Frontiers in earth science 7: 120. doi.org/10.3389/feart.2019.00120, 2019.

Zhang, Guang J., Xiaoliang Song, and Yong Wang. "The double ITCZ syndrome in GCMs: A coupled feedback problem among convection, clouds, atmospheric and ocean circulations." Atmospheric Research 229: 255-268. doi.org/10.1016/j.atmosres.2019.06.023, 2019.

Zhou, Lei, Raghu Murtugudde, Dake Chen, and Youmin Tang. "A Central Indian Ocean mode and heavy precipitation during the Indian summer monsoon." Journal of Climate 30, no. 6: 2055-2067. doi.org/10.1175/JCLI-D-16-0347.1, 2017.

Zhou, Y., Zaitchik, B. F., Kumar, S. V., Arsenault, K. R., Matin, M. A., Qamer, F. M., Zamora, R. A., and Shakya, K.: Developing a hydrological monitoring and sub-seasonal to seasonal forecasting system for South and Southeast Asian river basins, Hydrol. Earth Syst. Sci., 25, 41–61, https://doi.org/10.5194/hess-25-41-2021, 2021.

Zhou, Zhen-Qiang, Renhe Zhang, and Shang-Ping Xie. "Interannual variability of summer surface air temperature over central India: Implications for monsoon onset." Journal of Climate 32, no. 6: 1693-1706. doi.org/10.1175/JCLI-D-18-0675.1, 2019.

Figures

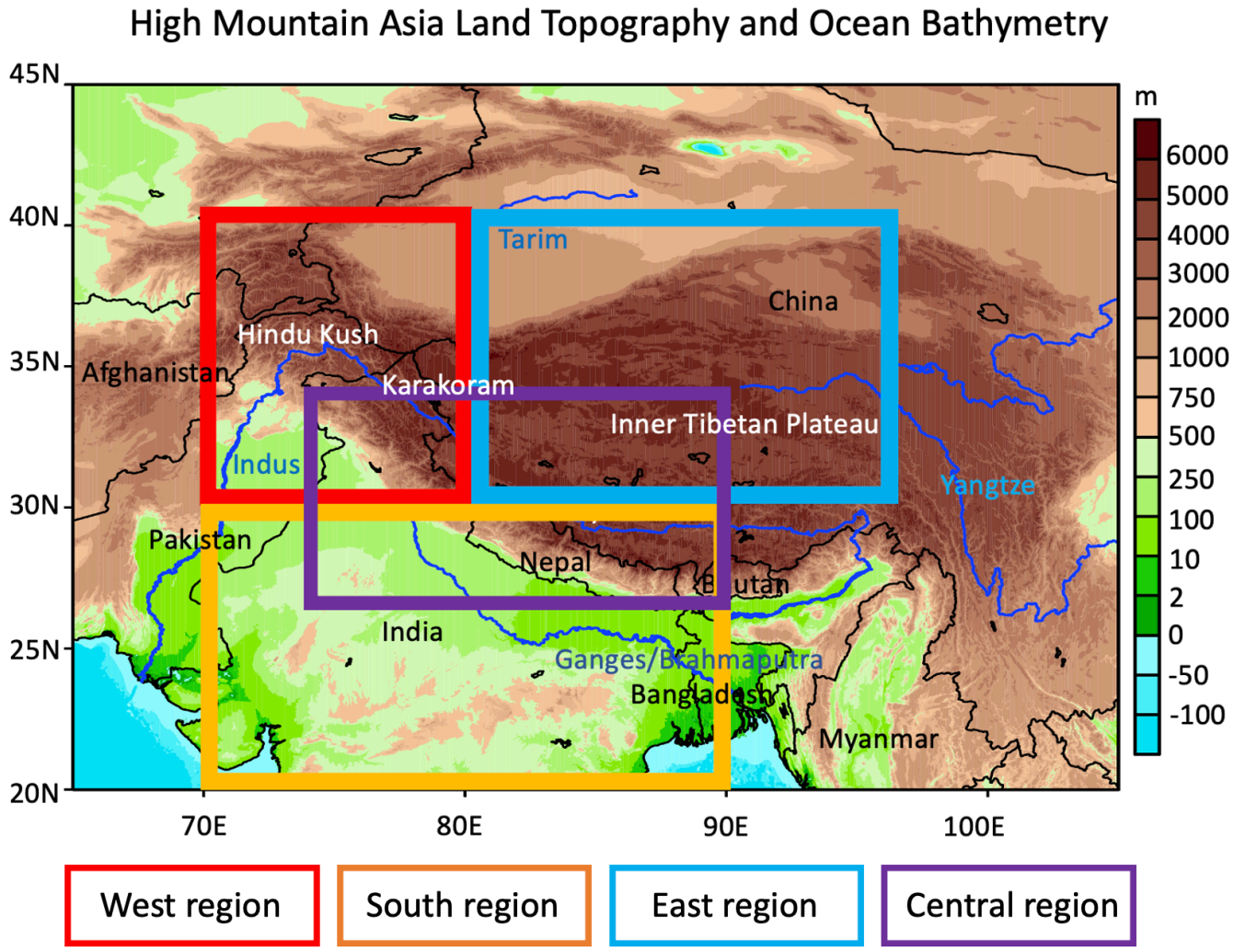

Figure 1: Topography and ocean bathymetry using the NOAA National Geophysical Data Center's ETOPO1 Global Relief Model. The map shows the elevation (m) for the HMA domain. The topography shown in this map is not the same as the topography used by GEOS-S2S, which has a coarser representation of the actual topography in the HMA region. In this figure, countries are in black text, mountain ranges in white text, and main rivers that are in major basins in blue text. Four subregions are defined for additional analysis, where the West region shown in red includes the Hindu Kush and Karakoram mountains, the South region includes the Indian subcontinent, the East region includes the Inner Tibet Plateau, and the Central region includes the Himalayas.

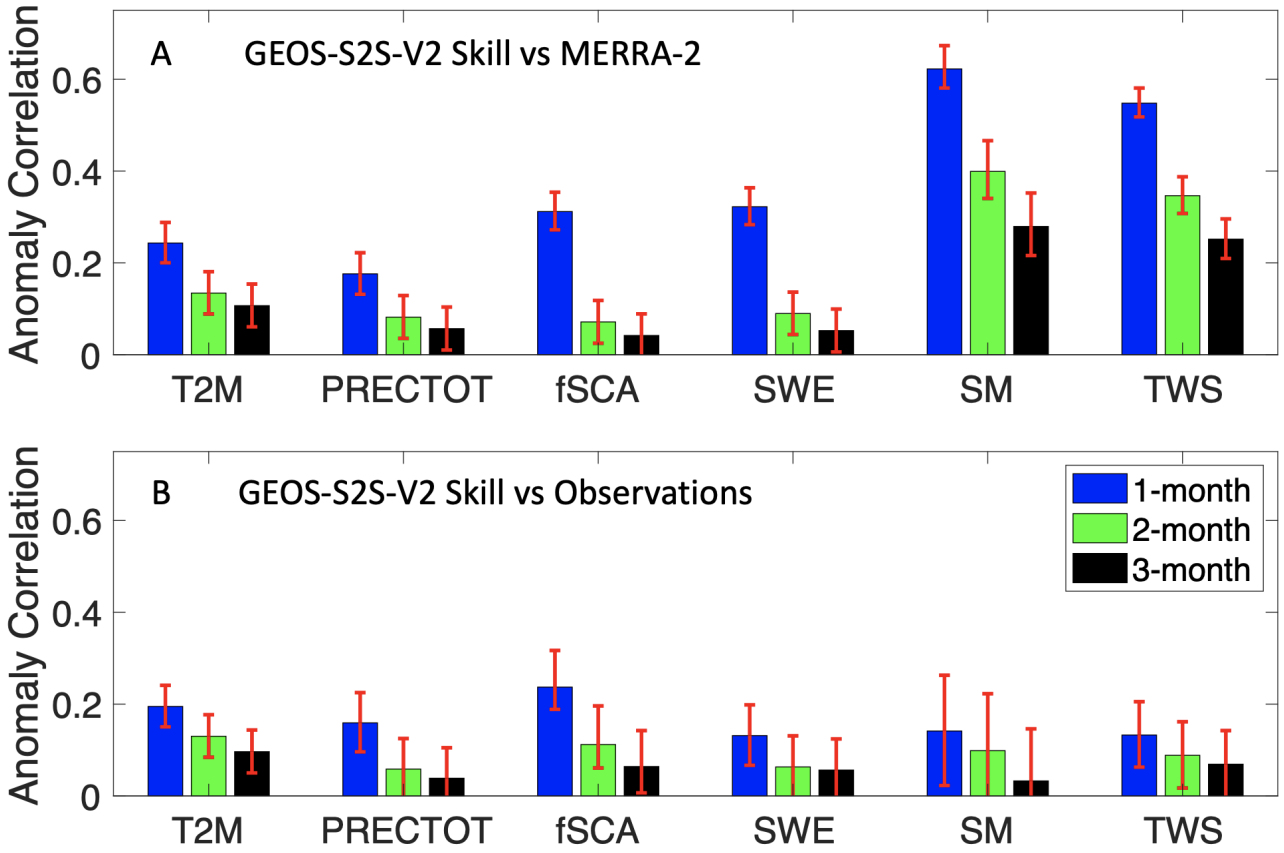

Figure 2: Anomaly correlation skill between variables for the GEOS-S2S forecasts when evaluated against MERRA-2 (Figure 2A) and against reference data products (Figure 2B). The evaluation of the 1-month lead forecasts is shown in the first bar (blue), 2-month in the second bar (green), and 3-month in the third bar (black). The red error bars indicate the spatial standard deviation of the anomaly correlation for each variable. The reference data that are used in Figure 2B are listed in Table 1.

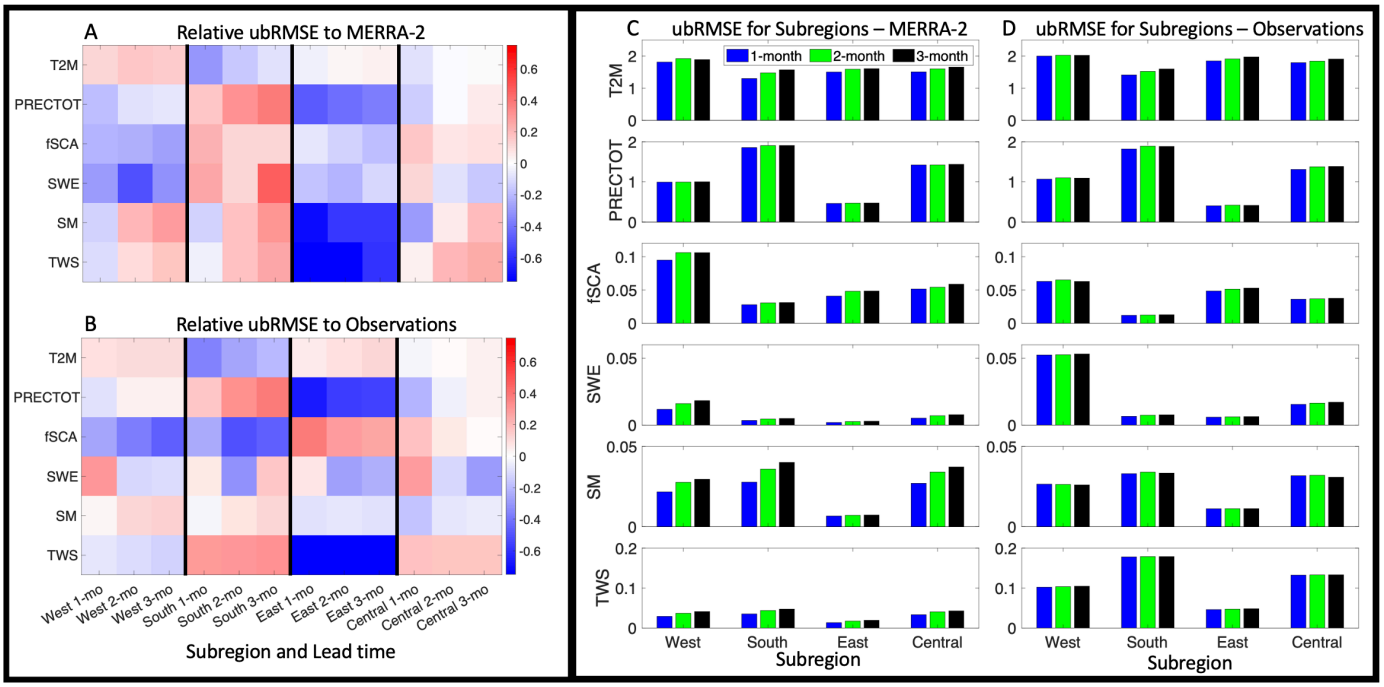

Figure 3: A) Portrait diagram visually depicting the S2S forecast skill when evaluated against MERRA-2, comparing the skill between the different subregions (e.g. West region 1-month forecast for T2M). Values in each box are the ubRMSE normalized by the absolute value of the climatological mean of that variable in that region (i.e. divided by the absolute value of mean T2M for the West region), normalized again by all the skill values for that climate variable (i.e. compare each metric with the normalized ubRMSE values of T2M for all subregions and all lead times). This means that if a box is blue (red), that climate variable in that subregion for that lead time has a lower (higher) normalized error when compared to that same climate variable in other subregions and lead times. B) Same as A) but when evaluating against other observations. C) Errors shown here are the original ubRMSE values for each subregion when evaluating against MERRA-2, shown for all lead times and separated by climate variable. D) Same as C) but when evaluating against other observations.

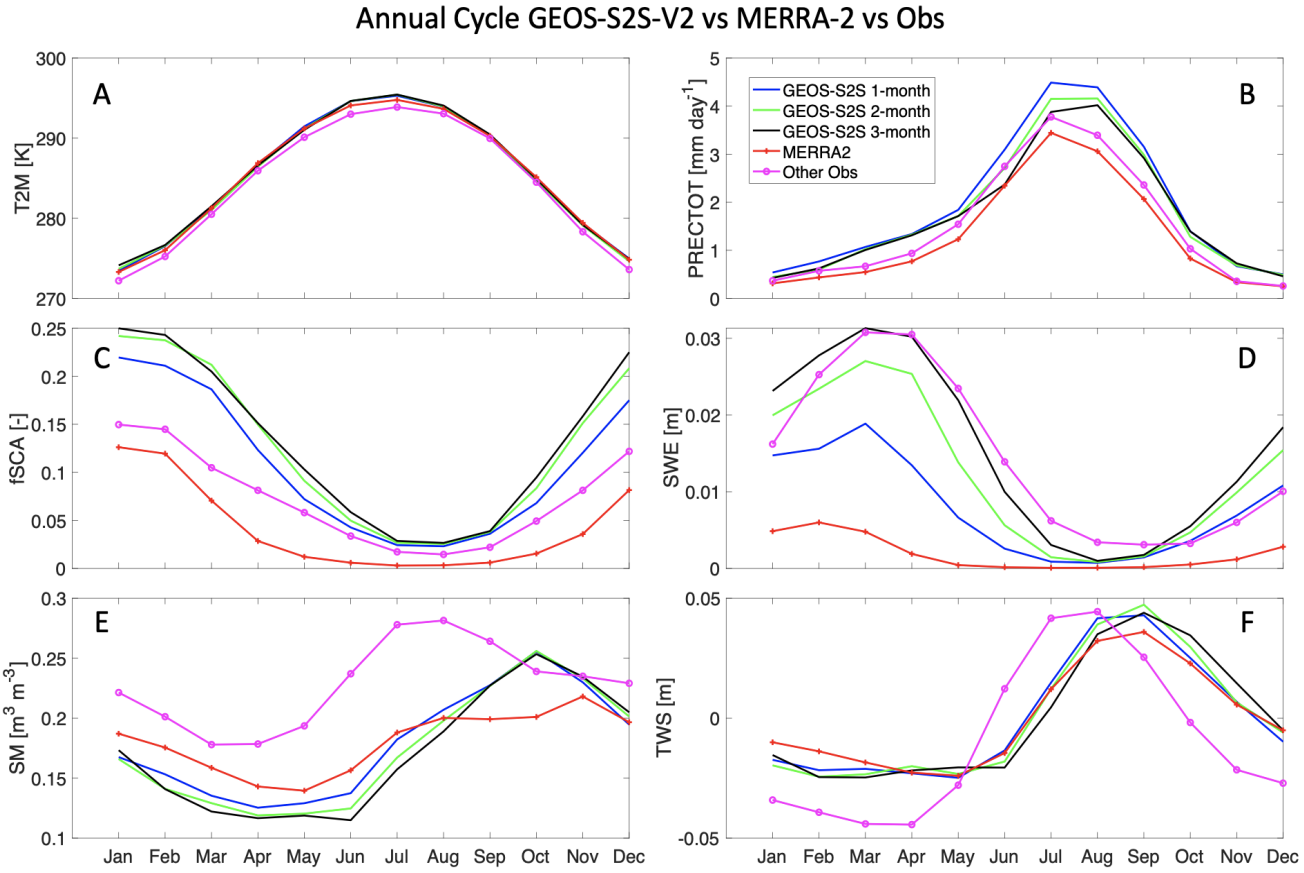

Figure 4: Annual cycle for each variable, averaged over the HMA domain. The annual cycles from the GEOS-S2S forecasts are shown for all lead times (blue, green, and black curves), and those estimated from the MERRA-2 reanalysis (red) and the reference data products (pink) are shown for comparison.

1205

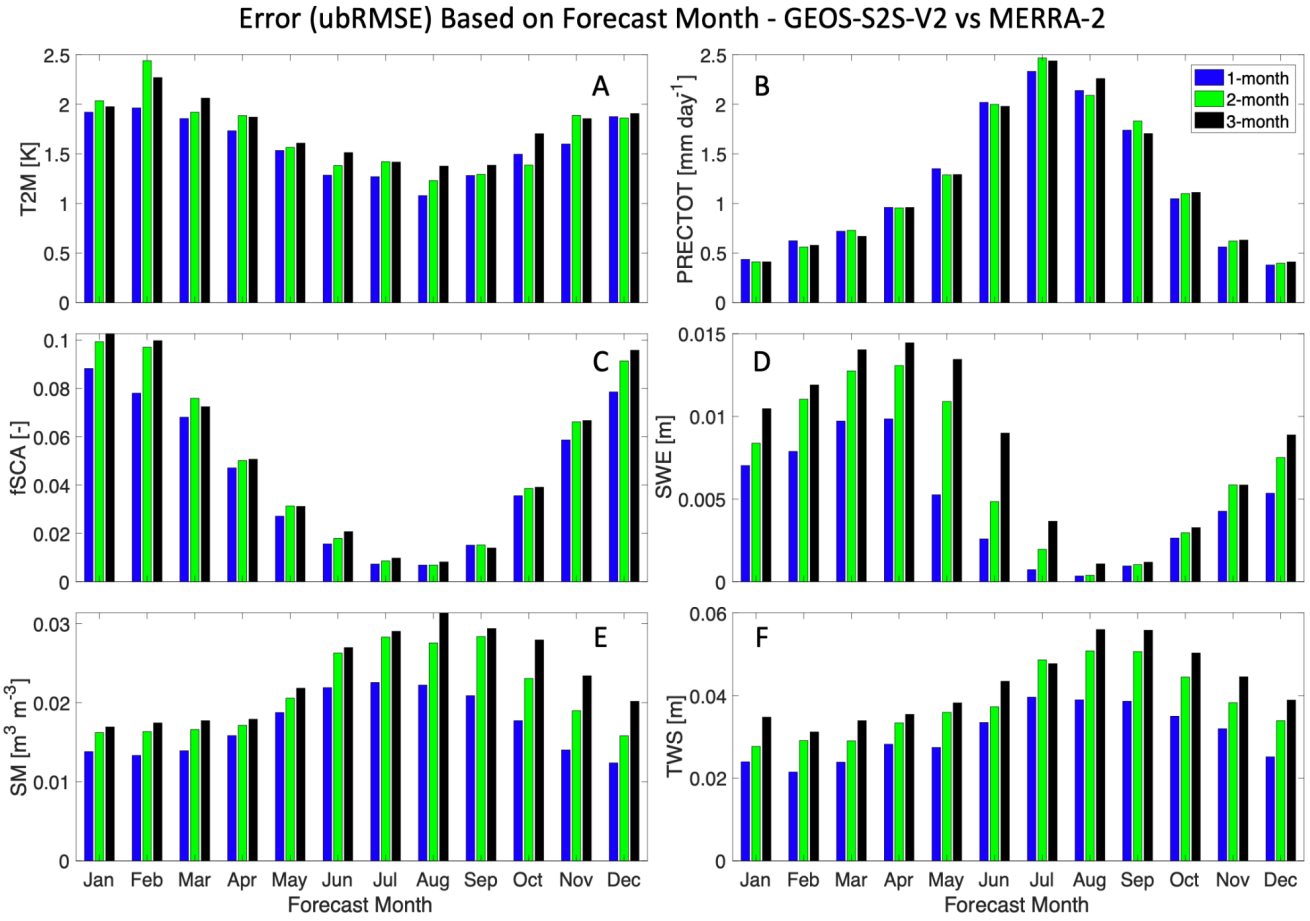

Figure 5: The expected error (ubRMSE) based on which month is forecasted. Shown here are results for 1-month (blue, first bar), 2-month (green, second bar), and 3-month (black, third bar) lead times for each variable. For example, the three bars for the month of April include the 1-month, 2-month, and 3-month forecasts, which are the forecasts initialized in March, February, and January respectively. The results displayed in this figure use MERRA-2 as the evaluation target.

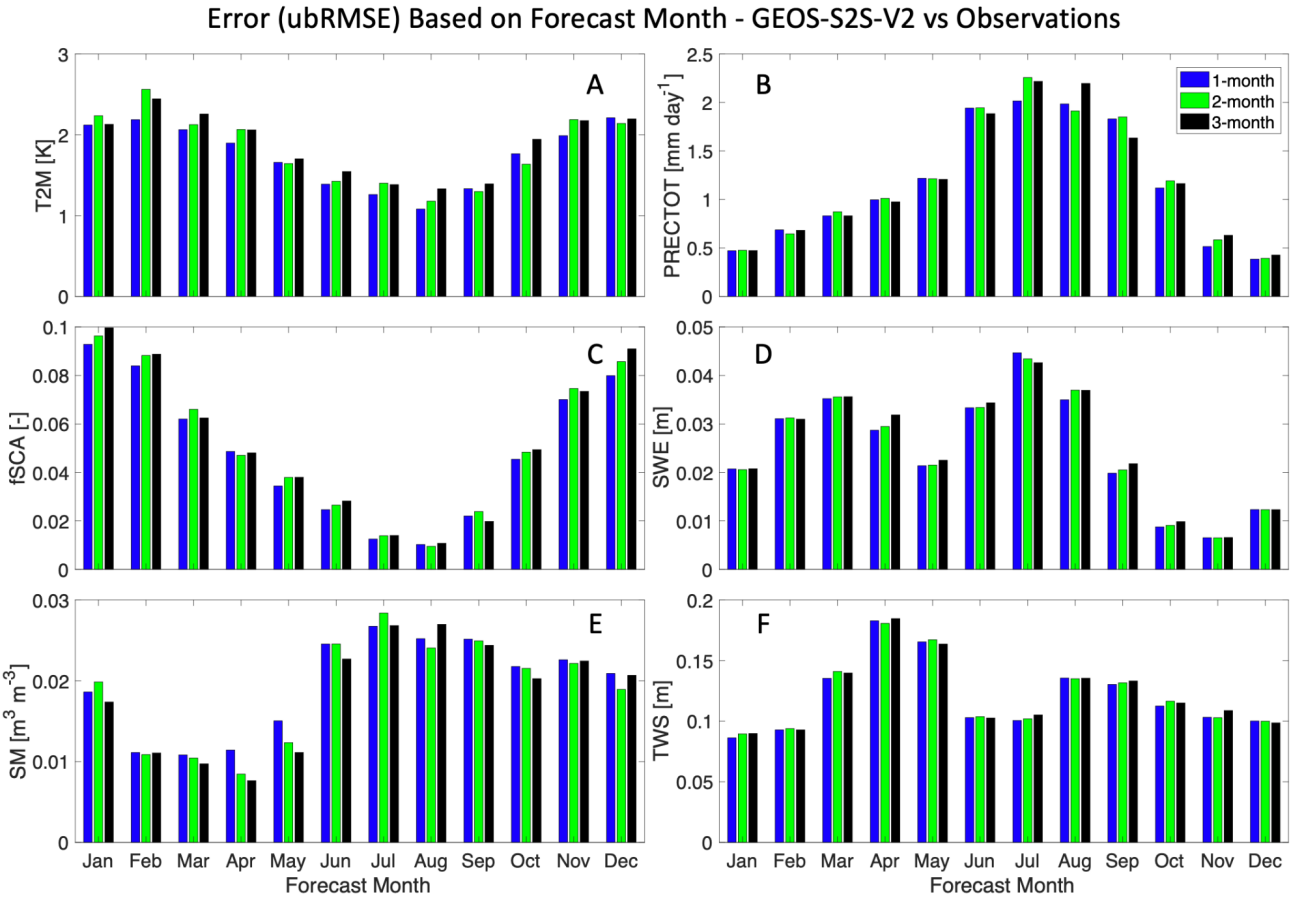

Figure 6: Same as Figure 5, but the results displayed in this figure use the reference data products as the evaluation target. The data that are used in this figure are: ERA5 for T2M, APHRODITE for PRECTOT, MODIS for fSCA, HMA-SR for SWE, ESA-CCI for SM, and GRACE for TWS.

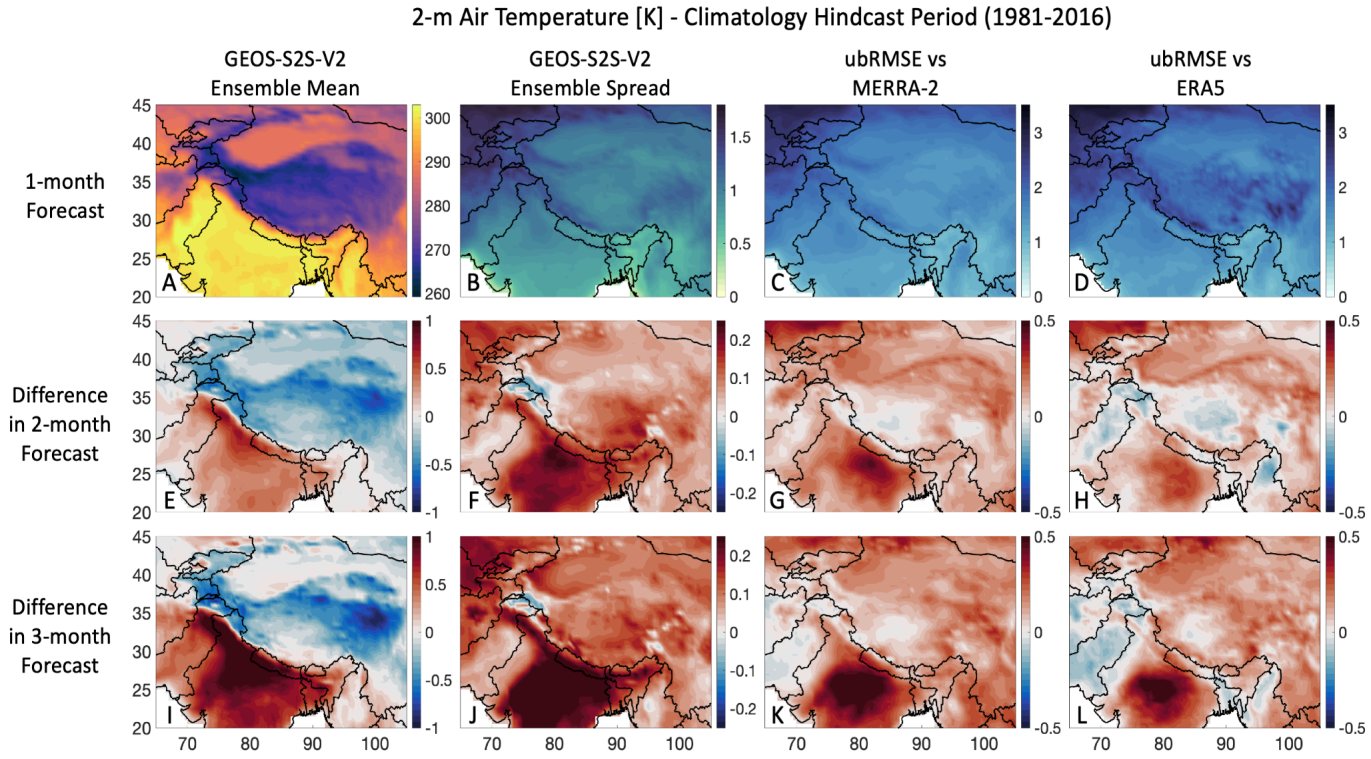

Figure 7: Screen-level (2-m) air temperature (T2M) metrics in HMA. Figure 7A shows the climatology (long term mean) from the GEOS-S2S 1-month forecast for the hindcast period (1981-2016). Figure 7B shows the ensemble spread from the GEOS-S2S 1-month forecast, calculated as the standard deviation of the model ensemble at each grid cell. Figures 7CD show the ubRMSE when comparing the GEOS-S2S 1-month forecast to MERRA-2 and ERA5, respectively. The bottom two rows of figures show the differences in the climatology, ensemble spread, and ubRMSE between the 2-month (Figures 7E-H) and 3-month (Figures 7I-L) forecasts compared to the 1-month forecast shown in the top row. Note, to calculate the difference shown in the bottom two rows, the 1-month maps in the top row are subtracted from the corresponding 2- and 3-month maps (i.e., 2-month maps minus 1-month maps and 3-month maps minus 1-month maps, respectively). Therefore, red in the subfigures indicates higher values (i.e., hotter temperature, larger spread, or larger error) in the 2- and 3-month forecasts, and blue indicates lower values compared to the 1-month forecasts. The units for these plots are in [K].

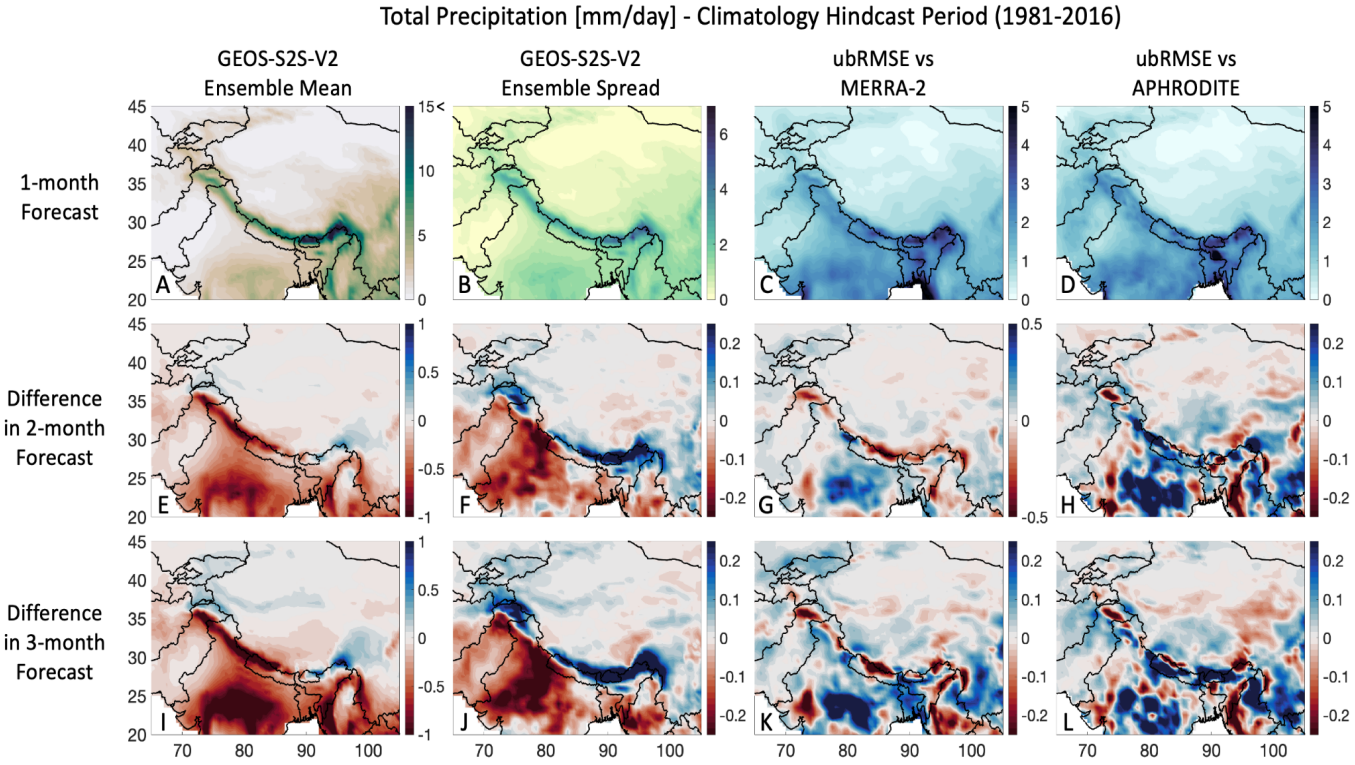

Figure 8: As in Figure 7, but for precipitation (PRECTOT) in [mm day^{-1}] and vs. APHRODITE in the right column. Here, red in the subfigures indicates lower values (i.e., less precipitation, smaller spread, or smaller error) in the 2- and 3-month forecasts and blue indicates higher values compared to the 1-month forecasts.

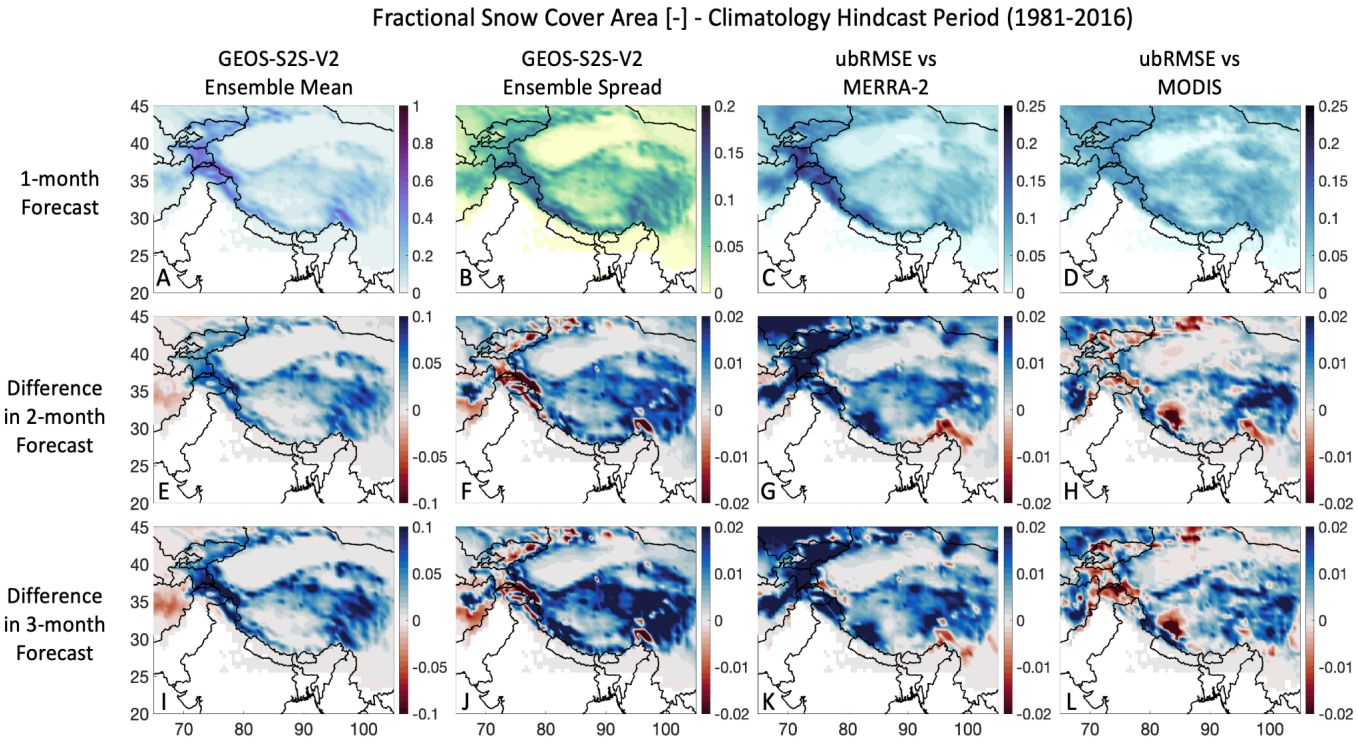

Figure 9: As in Figure 7, but for fractional snow cover area (fSCA) [unitless] and vs. MODIS in the right column. Grid cells that are masked out (in white) show areas with no-data-values. Here, red in the subfigures indicates lower values (i.e., less snow cover, smaller spread, or smaller error) in the 2- and 3-month forecasts and blue indicates higher values compared to the 1-month forecasts.

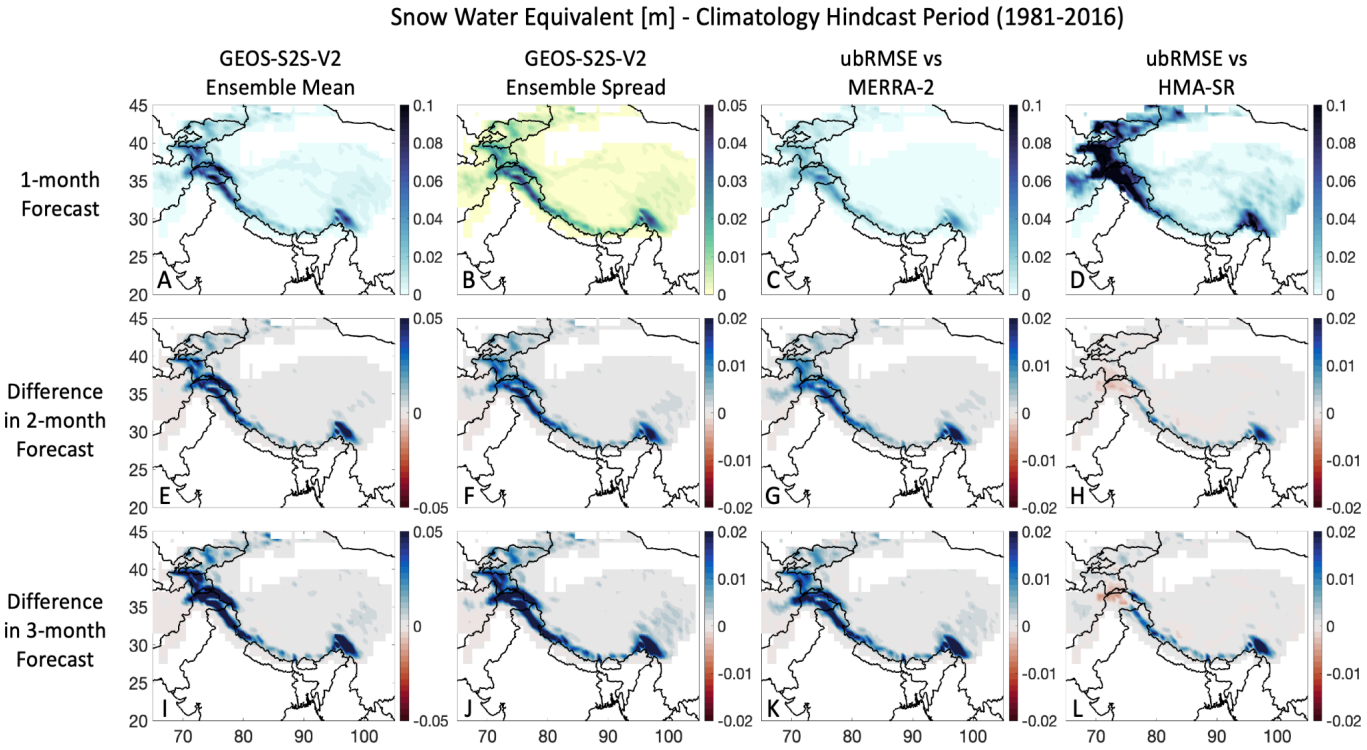

Figure 10: As in Figure 7, but for snow water equivalent (SWE) in [m] and vs. HMA-SR in the right column. Grid cells that are masked out (in white) show areas with no-data-values. Here, red in the subfigures indicates lower values (i.e., less snow water, smaller spread, or smaller error) in the 2- and 3-month forecasts and blue indicates higher values compared to the 1-month forecasts.

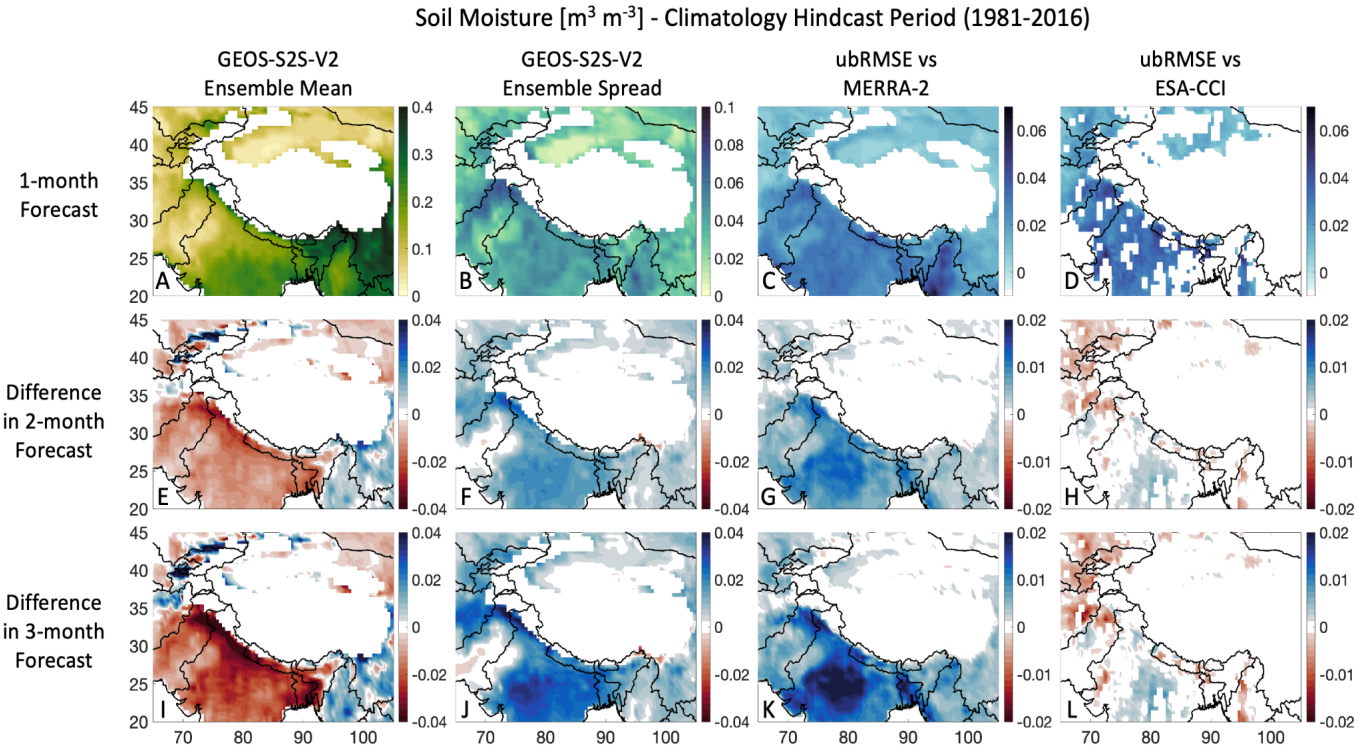

Figure 11: As in Figure 7, but for soil moisture (SM) in [$m^3 m^{-3}$] and vs. ESA-CCI in the right column. Grid cells that are masked out (in white) show areas with no-data-values. Here, red in the subfigures indicates lower values (i.e., less soil moisture, smaller spread, or smaller error) in the 2- and 3-month forecasts and blue indicates higher values compared to the 1-month forecasts.

250

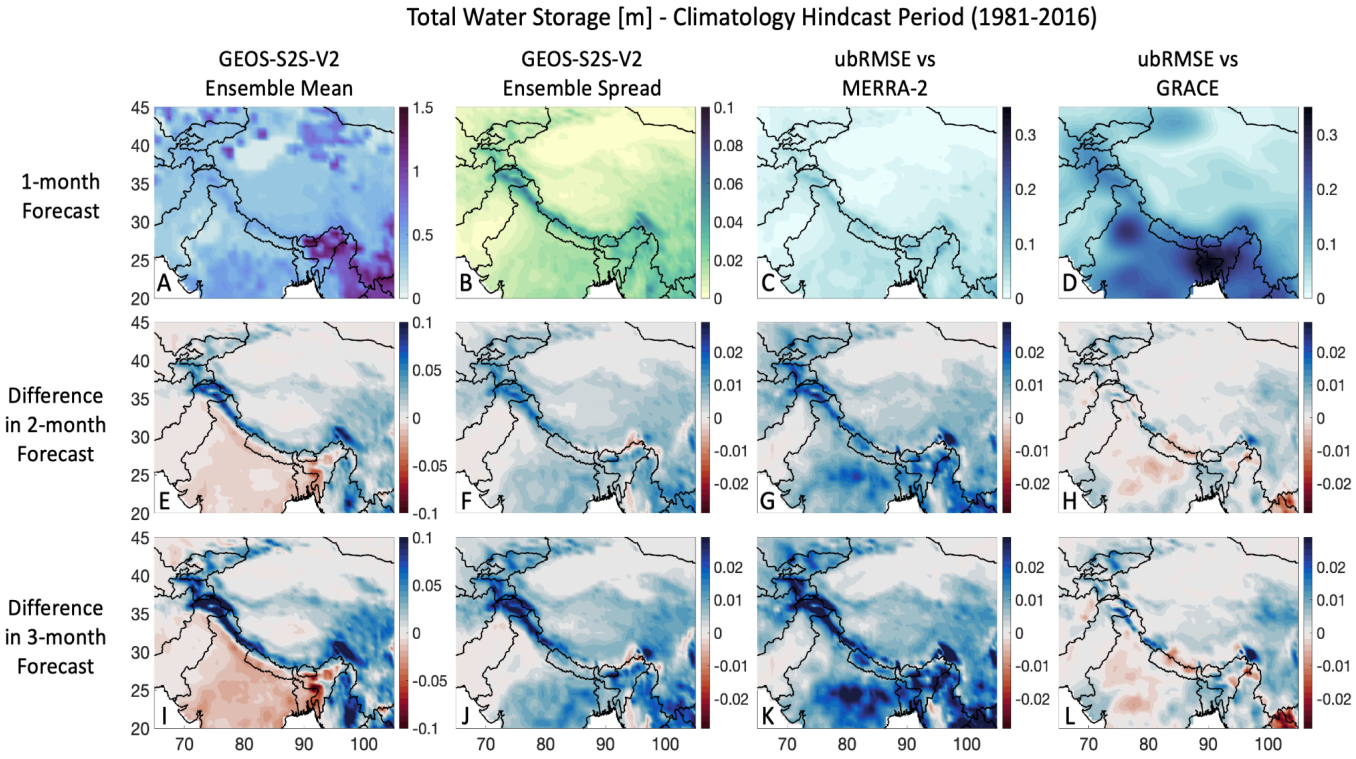

Figure 12: As in Figure 7, but for terrestrial water storage (TWS) in [m] and vs. GRACE in the right column. Here, red in the subfigures indicates lower values (i.e., less TWS, smaller spread, or smaller error) in the 2- and 3-month forecasts and blue indicates higher values compared to the 1-month forecasts.

1255

Tables

Table 1: The list of all data products used, including the GEOS-S2S V2 forecasting system, the MERRA-2 reanalysis product, and the various reference data products. The information in this table includes the period of data availability, the period used in the evaluation, the variables used in this study, the original spatial and temporal resolution, and the main reference for each data set. GEOS-S2S-V2, MERRA-2, and ERA5 data are provided up until the present day and production of these data sets occurs in near-real-time, where quality-assured monthly updates are typically published within 3 months of data production. GRACE data is originally provided at 3 degrees spatial resolution, but the version used here is posted at 1-degree spatial resolution.

Data Product	Available period	Evaluation period	Variables Evaluated	Spatial Resolution	Temporal Resolution	Reference
GEOS-S2S-V2	01/1981-12/2021	01/1981-12/2016	All variables	0.5 Degrees	Daily	Nakada et al., 2018
MERRA-2	01/1980-12/2021	01/1981-12/2016	All variables	0.625x0.5 Degrees	Hourly	Gelaro et al., 2017
ERA5	01/1979-12/2021	01/1981-12/2016	T2M	31 kilometers	3 hours	Hersbach et al., 2020
APHRODITE	01/1998-12/2015	01/1998-12/2015	PRECTOT	0.05 Degrees	Daily	Yatagai et al., 2012
MODIS	02/2000-12/2016	02/2000-12/2016	fSCA	0.05 Degrees	Daily	Hall et al., 2002
HMA-SR	10/1999-09/2017	01/2000-12/2016	SWE	500 meters	Daily	Liu et al., 2021b
ESA-CCI	01/1978-12/2020	01/2000-12/2016	SM	0.25 Degrees	Daily	Dorigo et al., 2017
GRACE	04/2002-10/2017	04/2002-12/2016	TWS	3 Degrees	Monthly	Tapley et al., 2004

Table 2: The unbiased RMSE (ubRMSE) and the anomaly correlation (Ranom) for all variables when comparing the GEOS-S2S forecasts to the reanalysis ('MERRA-2') and the reference data products ('Reference data'). The reference data that are used here are: ERA5 for T2M, APHRODITE for PRECTOT, MODIS for fSCA, HMA-SR for SWE, ESA-CCI for SM, and GRACE for TWS (Section 2.3).

		ubRMSE			**Ranom**	
GEOS-S2S vs MERRA2	**1-month**	**2-month**	**3-month**	**1-month**	**2-month**	**3-month**
T2M [K]	1.61	1.74	1.77	0.24	0.13	0.11
PRECTOT [mm day^{-1}]	1.06	1.08	1.08	0.18	0.08	0.06
fSCA [-]	0.035	0.041	0.041	0.31	0.07	0.04
SWE [m]	0.002	0.003	0.004	0.32	0.09	0.05
SM [m^3 m^{-3}]	0.019	0.023	0.025	0.62	0.40	0.28
TWS [m]	0.025	0.032	0.035	0.55	0.35	0.25
		ubRMSE			**Ranom**	
GEOS-S2S vs Reference data	**1-month**	**2-month**	**3-month**	**1-month**	**2-month**	**3-month**
T2M [K]	1.78	1.87	1.90	0.19	0.13	0.10
PRECTOT [mm day^{-1}]	1.03	1.06	1.06	0.16	0.06	0.04
fSCA [-]	0.048	0.051	0.052	0.24	0.11	0.06
SWE [m]	0.021	0.021	0.022	0.13	0.06	0.06
SM [m^3 m^{-3}]	0.021	0.020	0.020	0.14	0.10	0.03
TWS [m]	0.101	0.103	0.104	0.13	0.09	0.07

Supplementary Information

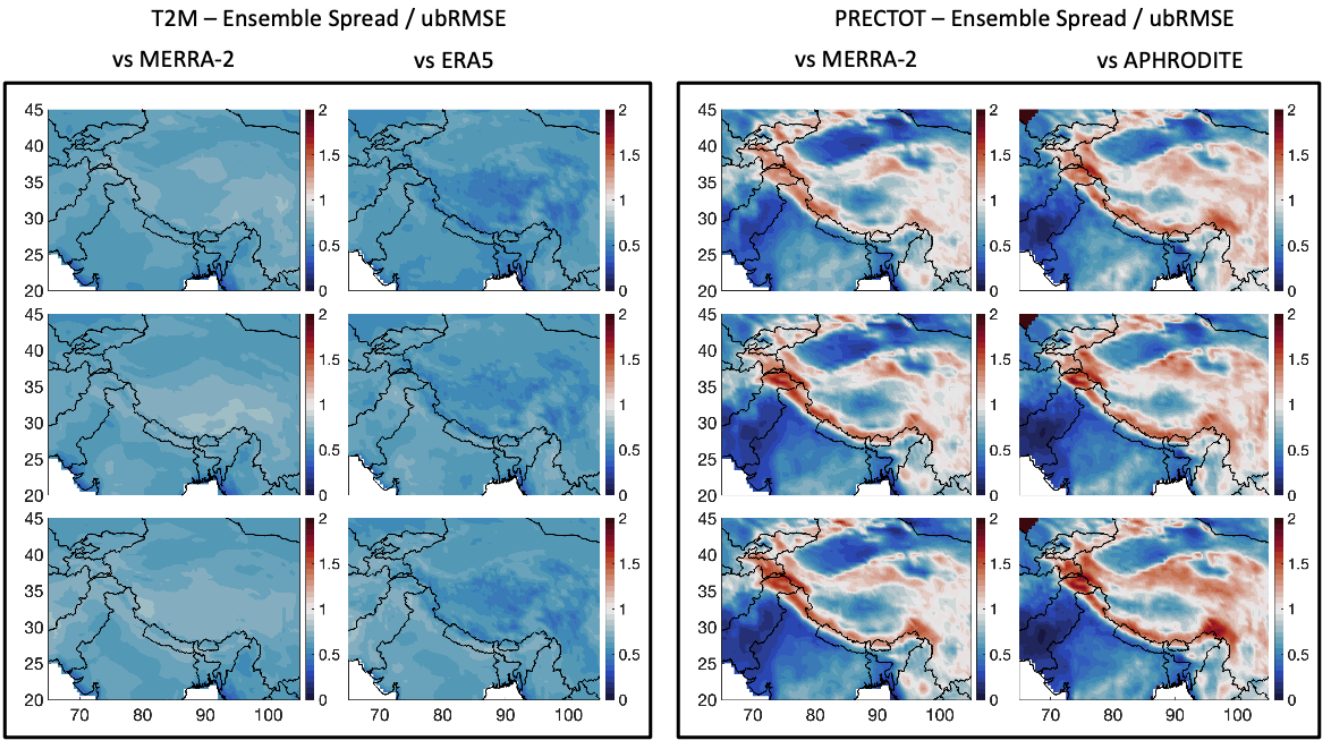

Figure S1: These plots show the reliability of the S2S forecasts for T2M and PRECTOT, calculated as the ensemble spread divided by the ubRMSE, information that can be found in Figures 7-8. For these plots, blue indicates the forecasts are overconfident for that region, meaning there is a smaller spread compared to what the error is. Red shows the opposite, indicating that the forecasts are underconfident for that region, which means there is a larger spread compared to what the error is. The rows of plots show the 1-month (top), 2-month (center), and 3-month (bottom) forecasts.

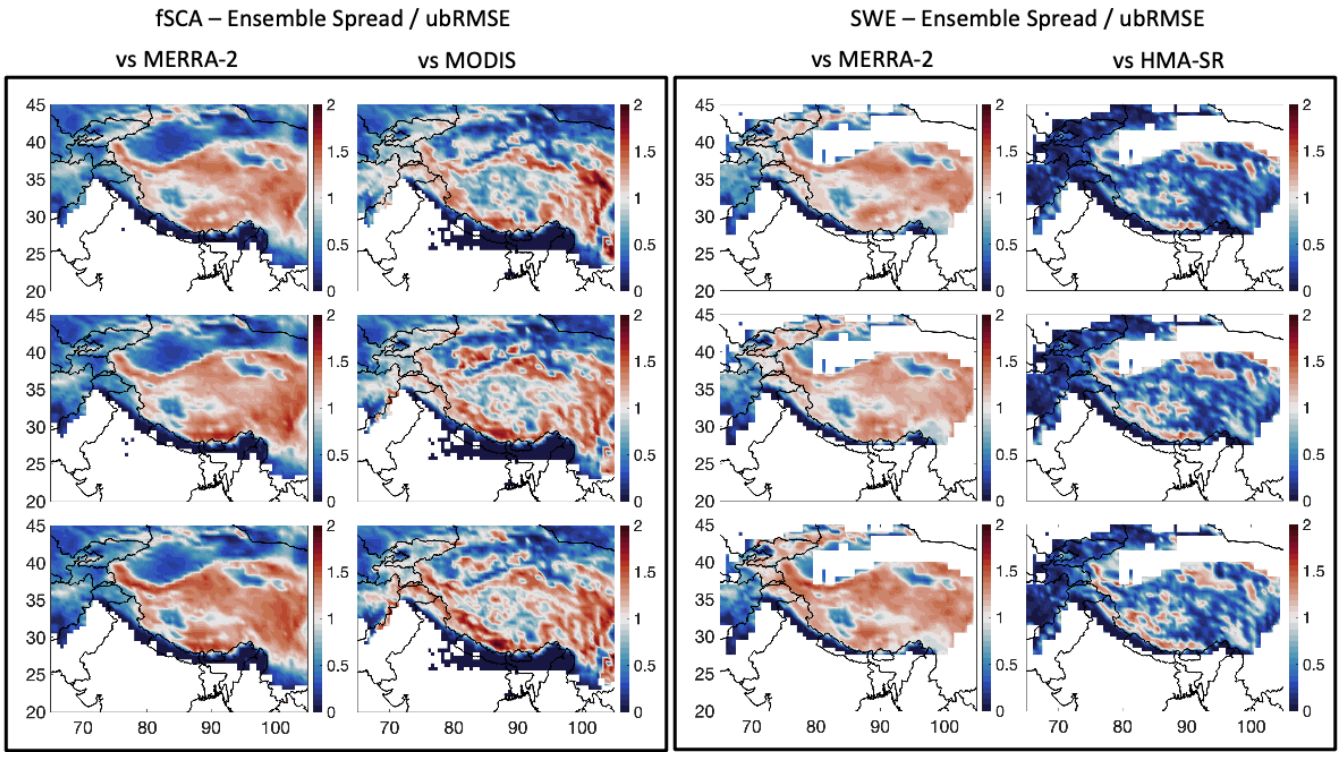

Figure S2: These plots show the reliability of the S2S forecasts for fSCA and SWE, calculated as the ensemble spread divided by the ubRMSE, information that can be found in Figures 9-10. For these plots, blue indicates the forecasts are overconfident for that region, meaning there is a smaller spread compared to what the error is. Red shows the opposite, indicating that the forecasts are underconfident for that region, which means there is a larger spread compared to what the error is. The rows of plots show the 1-month (top), 2-month (center), and 3-month (bottom) forecasts.

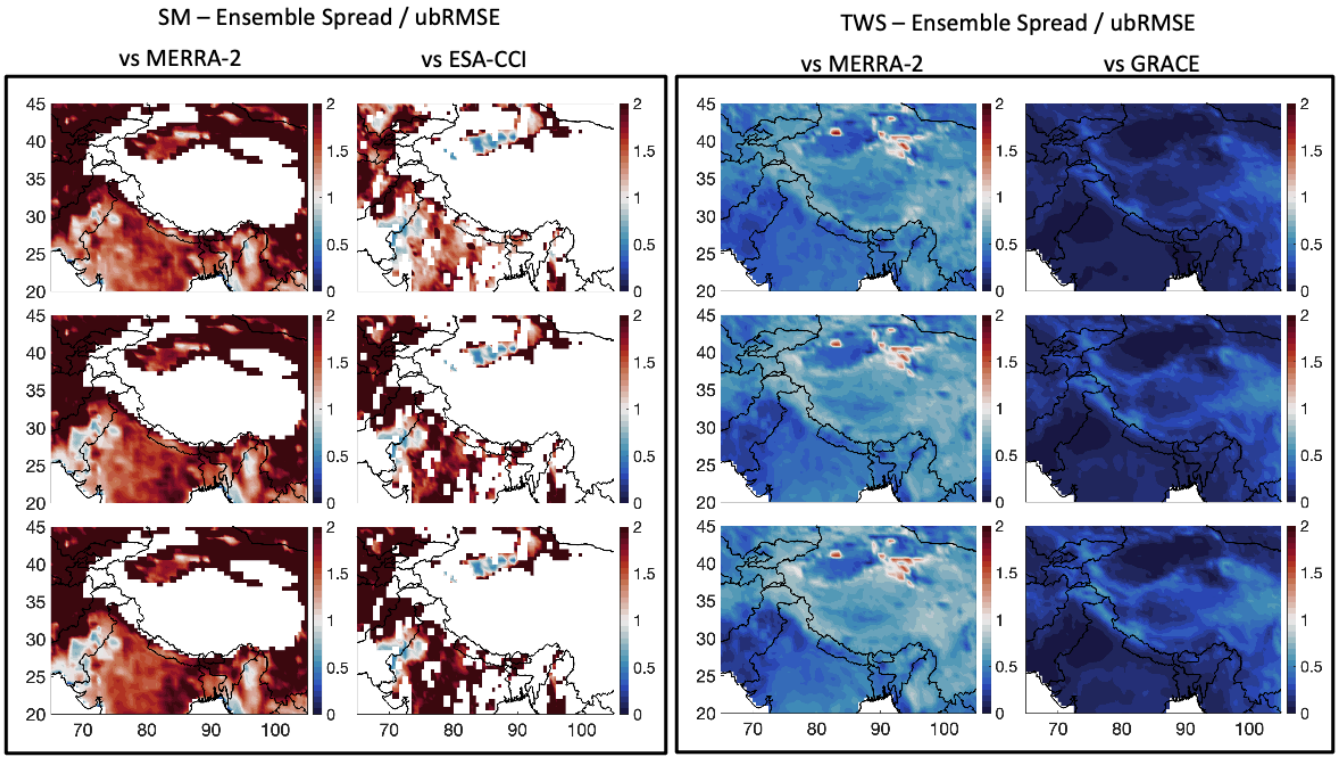

Figure S3: These plots show the reliability of the S2S forecasts for SM and TWS, calculated as the ensemble spread divided by the ubRMSE, information that can be found in Figures 11-12. For these plots, blue indicates the forecasts are overconfident for that region, meaning there is a smaller spread compared to what the error is. Red shows the opposite, indicating that the forecasts are underconfident for that region, which means there is a larger spread compared to what the error is. The rows of plots show the 1-month (top), 2-month (center), and 3-month (bottom) forecasts.