UTRECHT UNIVERSITY

Department of Information and Computing Science

Applied Data Science master thesis

Using satellite products to improve random forest-based streamflow simulations from a global hydrological model

First examiner:
Derek Karssenberg (UU)

Candidate:
Niek Collot d'Escury (6256481)

Second examiners:
Edwin Sutanudjaja (UU)
Michele Magni (UU)
Oriol Pomarol Moya (UU)

§§

June 30, 2023

Abstract

Hydrological models are important tools for making streamflow predictions and studying the effects of climate change on water resources around the world. This study aims to improve a global hybrid hydrological streamflow framework based on PCR-GLOBWB and Random Forest (RF), with the addition of satellite products. The selected satellite products are liquid water equivalent from GRACE (Gravity Recovery and Climate Experiment), snow cover fraction from JASMES (JAXA Satellite Monitoring for Environmental Studies) and soil moisture from ESA CCI SM (European Space Agency Climate Change Initiative Soil Moisture). These products are selected because of their relevance for streamflow predictions. Five global and seven Local (Australia, Canada and the United States) model configurations were used for the RF model. The differences between configurations are based on the inclusion of state variables from PCR-GLOBWB, catchment attributes, satellite products, lagged variables for meteorological input and satellite products of 4 and 12 months and exclusion of state variables from PCR-GLOBWB related to satellite products. The results showed that the global and local configurations did not improve compared to the benchmark model (based on 51 predictors from PCR-GLOBWB) and the addition of lag was not significantly effective either. The configuration with only satellite and meteorological input and satellite products did, however show good performance with only six predictors. After adding static variables to the previous configuration there was equal performance to configurations with state variables from PCR-GLOBWB. These result mean that good performance can be achieved without the need for a hydrological model and with a limited number of variables.

Contents

1 Introduction **4**

2 Data **7**
2.1 Pre-processing 8
2.2 PCR-GLOBWB 8
2.3 GRACE 9
2.4 SCF JASMES 10
2.5 ESA CCI SM 11

3 Method **12**
3.1 Random Forest 12
3.2 Model setup 13

4 Results **17**
4.1 Global runs 17
4.2 Local runs 24

5 Discussion **26**
5.1 Limitations and recommendations 27

6 Conclusion **28**

Appendix

A Appendix **29**

B Appendix **30**

C Appendix **32**

D Appendix **33**
D.1 Variable importance local configurations 33
D.2 Boxplots local configurations 36

Bibliography **44**

1. Introduction

Accurate hydrological models are a very import aspect of managing the impact of climate change on water resources around the world [1]. The modelling of the global hydrological cycle is done with Global Hydrological Models (GHM), which simulate hydrological responses to weather and climate variations. GHM's are used for a variety of applications, including flood and drought forecasting, studying the impact of climate change and disaster management [2], [3]. All hydrological models contain errors due to different factors, such as model structure, parameter values, initial conditions or observational input data [4]. Traditionally, these errors have been tackled with optimization techniques that aim at calibrating model parameters to better fit historical streamflow records [5]–[7]. Recently, however, different studies have investigated the use of hybrid global streamflow modeling frameworks, where statistical learning (SL) is used to improve GHM's [8]–[10]. Many different SL algorithms can be used to improve streamflow predictions [11].

Shen *et al.* (2022)[8] developed a hybrid global streamflow framework using a Random Fores (RF) based error updating procedure to improve streamflow predictions from the PCR-GLOBWB GHM for three stations in the Rhine and Meuse basins. The study used an extensive set of hydrological variables obtained from PCR-GLOBWB and meteorological input variables (precipitation, temperature and reference potential evapotranspiration). The dependent variable for the RF model was the difference (error) between the observed streamflow and the modeled streamflow by PCR-GLOBWB. Interestingly they found that the RF model performed better for an uncalibrated PCR-GLOBWB. Magni *et al.* (2023)[10] extended this work to a global scale, increased the number of predictors obtained from PCR-GLOBWB for the RF model, changed the dependent variable from an error to the observed streamflow and included the modeled streamflow as a predictor. They achieved good performance for most stations and significantly improved the predictions made by PCR-GLOBWB alone.

The use of a hybrid framework, as used by Magni *et al.* (2023)[10] does

not solve the problem of incorrect parameter estimation [12]. Satellite-based remote sensing, however, can provide an alternative to estimated hydrological state variables, such as soil moisture, precipitation or snow water equivalent [12]. The type of variables that can be estimated with remote sensing raise the extra question of the use of lagged variables. Since all three variables have lagged effects in the hydrological cycle [13]–[15].

The aim of this study is to test if the use of satellite products can improve the proposed hybrid framework of Magni *et al.* (2023)[10]. Satellite data can be used as input forcing for a hydrological model or as input for the post-processor of the hybrid framework. In this study we focus on the use of satellite data as input for the post-processor. Although the use of satellite data is not fit for long-term forecasting, it can be used for near-real time forecasting [16]–[19] or to create a global streamflow reanalysis dataset [20]. This type of dataset is used for improvement of modeled streamflow values in ungauged basins [21], [22] and becomes increasingly valuable due to the trend in the reduction of the number of global gauging stations [23].

The main question of this study is:

How can the performance of a hybrid streamflow model be improved by replacing specific state variables of the PCR-GLOBWB model with satellite data?

To assist in addressing the main question, the following three sub-questions are employed.

- Is it possible to largely or completely remove the PCR-GLOBWB variables as input for the post processor, while retaining high performance?
- Does lag added to the satellite and meteorological variables improve performance?
- Is the hypothesized improvement dependent on region?

In order to answer the stated research question, the initial focus will be data availability. Before satellite data can be used in the framework of Magni *et al.* (2023)[10], station values needed to be extracted for all catchments in the study. The RF model was trained with five global and seven local configurations. The configurations vary on the inclusion of PCR-GLOBWB state variables, catchment attributes and lagged satellite or meteorological vari-

ables. The local configurations are used to investigate if the performance depends on the region and these are used to evaluate the lagged configurations. The global configurations are not suitable for evaluating the lagged configurations, since the large number of variables makes these configurations computationally unfeasible.

This thesis is organized as follows. In section 2 an overview of the data used is presented. Subsequently, section 3 gives a detailed description of the methodology of how the satellite data is incorporated in the work of Magni *et al.* (2023)[10] followed by the results in section 4. Section 5 presents discussion points, while conclusions are given in section 6.

2. Data

Data for this study is obtained from three different satellite products and from previous research. Data from Magni *et al.* (2023)[10] can be found at Zenodo [24] and contains streamflow observations, meteorological variables and fifty-one features from PCR-GLOWB for a period of 1979 to 2019 (a complete overview of all included features can be found in Table B.1). The meteorological variables precipitation, temperature and reference potential evapotranspiration were obtained from the CRU TS 3.2 dataset [25]. The streamflow data is available on the Global Runoff Data Centre (GRDC). The station selection criteria were a minimum upstream catchment area of 10.000 km^2 (based on Magni *et al.* (2023)[10]) and at least one month of observations without missing values for the period of 2002 and 2019. Figure 1 gives an overview of the available streamflow data for the 1227 selected stations. We focus on the period Feb-2002 until Dec-2019, as most of the satellite products are available for this period. More information about the PCR-GLOBWB model can be found in section 2.2. Sections 2.3, 2.4 and 2.5 introduce the selected satellite products.

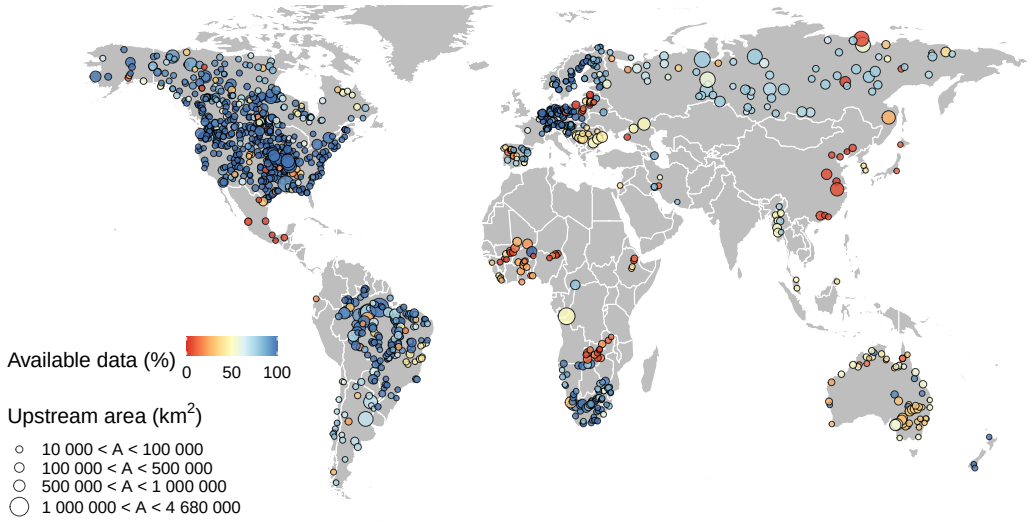

Figure 1: Global coverage of streamflow data. The color scale indicates the percentage of missing discharge values for the period 2002-2019 and the size of the point indicates the size of the catchment area for each station.

2.1 Pre-processing

Data made available by Magni *et al.* (2023)[10] was already pre-processed and required no further manipulation. The satellite data, however, required the same upstream normalization as performed on PCR-GLOBWB inputs and outputs. Before the normalization could be achieved all three satellite products were first transformed to the same 0.5^o resolution (using the setgrid cdo command [26]). Afterwards the upstream normalization was achieved with the PCRaster Python framework [27]. Station values were extracted by locating the nearest pixel corresponding to the GRDC station coordinates. Any missing cell values in the upstream area of a station causes the calculation for the entire catchment to result in a missing values. By changing the missing cell value in the upstream area to zero prior to the upstream normalization, the number of missing values per station was significantly reduced, since the entire catchment needs to be missing before the calculation results in a missing value. All the stations that did not contain at least one month of data without missing values in either the observations or the satellite data was excluded from the study.

To investigate the effect of lagged variables in this study, lag was added to the satellite and meteorological variables. Four and twelve months lag were selected, due to the potential inclusion of seasonal or yearly lagged patterns.

2.2 PCR-GLOBWB

PCR-GLOBWB is a grid-based GHM that can be run with different spatial resolutions. This study focuses on a uncalibrated PCR-GLOBWB run with a 30 arcmin (±50 km resolution at the equator) resolution [28]. PCR-GLOBWB takes into account both natural hydrological processes as well as anthropogenic influences (see various withdrawal features in Table B.1) [29]–[31].

PCR-GLOBWB was run with a daily timestep, after which the outputs were upscaled to a monthly timestep to reduce computation time and make them compatible with the satellite products used in this study.

2.3 GRACE

The Gravity Recovery and Climate Experiment (GRACE) mission was a joint operation between the American National Aeronautics and Space Administration and the German aerospace center starting in early-2002, with the purpose of creating a precise survey of Earth's time-variable gravity field [32]. The mission consisted of two identical satellites, that employed a satellite-to-satellite tracking system. The time varying-gravity field is measured by calculating the integrated difference in gravity acceleration of each satellite [33]. Known gravity effects such as tides and and non-tidal mass variations are removed as a post-processing step, which leaves the hydrological signal due to water mass variations on land. There are, however, still signals left from other processes including post-glacial rebound and seismic activity [34]. From the final measured gravity field the changes in the Terrestrial Water Storage (TWS) can be calculated, which will be referred to as the Liquid Water Equivalent (LWE) throughout this study.

Figure 2 gives an overview of the data availability for each GRDC station, showing overall a good global coverage. The 15 year period contains 24 months of missing data.

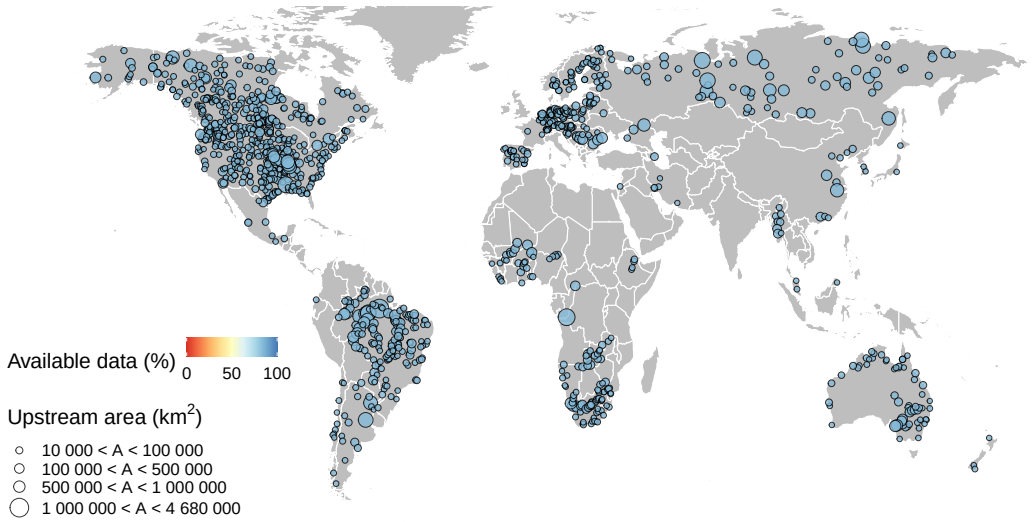

Figure 2: Global coverage of LWE. The color scale indicates the percentage of missing LWE values for the period 2002-2019 and the size of the point indicates the size of the catchment area for each station.

2.4 SCF JASMES

The Snow Cover Fraction (SCF) product of the JAXA Satellite Monitoring for Environmental Studies (JASMES) used in this study is derived from radiance data obtained from two different satellite series: the Advanced Very High Resolution Radiometer (AVHRR) and Moderate Resolution Imaging Spectroradiometer (MODIS) [35].

The AVHRR is used to obtain data from 1978 until 2005 and the MODIS instruments are used to get data from 2000 till 2021 (originally until 2015, which was later extended to 2021). The dataset has a spatial resolution of approximately 5 km and a bi-monthly temporal resolution. The radiance data of the two satellite series was used to create a snow cover extent dataset, which was converted to SCF [35]. The snow fraction values range from zero (no snow) to one (completely covered).

Figure 3 gives an overview of the data availability for each station. Most stations have a complete record of observations for SCF and there is no spatial pattern to the missing values. The maximum percentage of missing values for a single station is less than one percent.

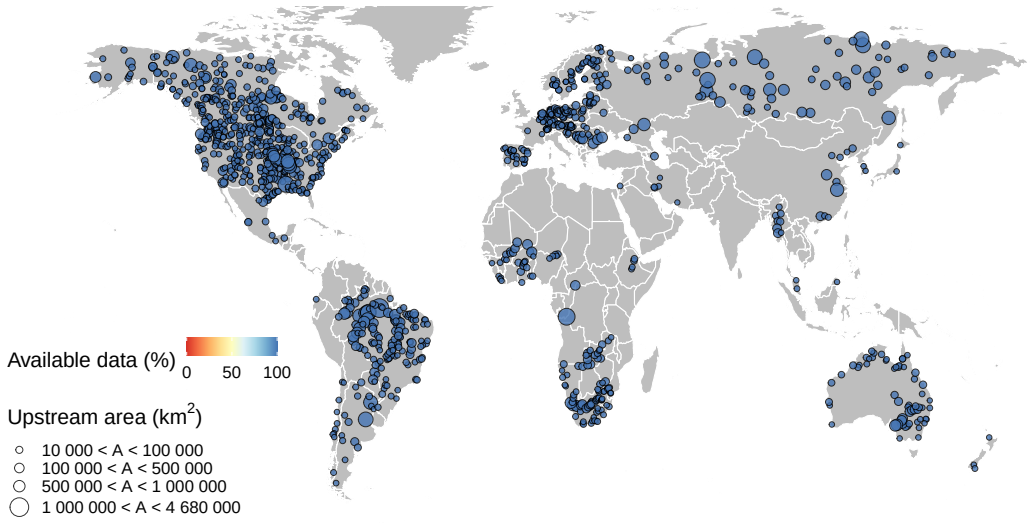

Figure 3: Global coverage of SCF. The color scale indicates the percentage of missing SCF values for the period 2002-2019 and the size of the point indicates the size of the catchment area for each station.

2.5 ESA CCI SM

The European Space Agency Climate Change Initiative Soil Moisture (ESA CCI SM) is a remote sensing product, where many different satellite products are combined to form a single Soil Moisture (SM) dataset. Before a satellite is included in this SM product it needs to satisfy certain criteria, to ensure that the required quality of the overall product is not diminished [36]. All satellite products included in the dataset are generated with microwave remote sensing tools to measure soil moisture. The current dataset covers a period of 1979 to 2019 with a spatial resolution of 0.5^{o} degrees. The dataset is regularly updated, changing either the temporal or spatial resolution and extent [36].

Figure 4 gives an overview of the data availability for each GRDC station. There is a clear pattern that is caused by limitations of the measuring equipment. Microwave remote sensing tools are not capable of accurately measuring SM in areas with abundant snow or ice cover [37]. This means there is an increase in missing data during winter. Dense vegetation, like in the tropics, is another cause for inadequate observations [38]. Stations where the entire record was missing for this reason were entirely removed from the current study.

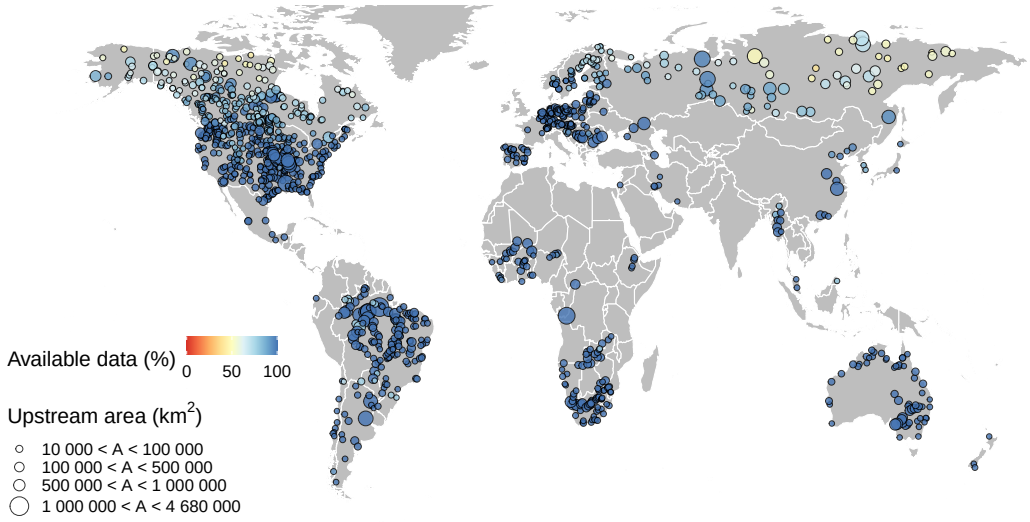

Figure 4: Global coverage of SM. The color scale indicates the percentage of missing SM values for the period 2002-2019 and the size of the point indicates the size of the catchment area for each station.

3. Method

In this study, a similar hybrid modelling approach is used as in Magni *et al.* (2023)[10], with different combinations of input variables for the SL algorithm. The choice of the correct satellite products that may help improve streamflow predictions depend on knowledge of the hydrological cycle [39]. The terrestrial water storage (TWS) is a crucial part of the hydrological cycle and can be calculated by summing up different parts of the water cycle [39], [40]:

$$TWS = SnWS + CWS + SWS + SMS + GWS \tag{1}$$

In Equation 1 the different storage components represent: Snow Water Storage ($SnWS$), Canopy Water Storage (CWS), Surface Water Storage (SWS), Soil Moisture Storage (SMS) and Ground Water Storage (GWS). Liquid Water Storage (LWE) from GRACE was selected to replace TWS, Soil Moisture (SM) from ESA CCI SM to replace SMS and Snow Cover Fraction (SCF) from JASMES to serve as proxy for $SnWS$.

In the following sections the RF algorithm (section 3.1) and model setup (section 3.2 will be further explained.

3.1 Random Forest

RF is a frequently used ensemble learning algorithm that uses multiple decision trees as base learners and is an extension of bagging classification trees [41]. To reach a final result, a tree is developed by minimizing an error function for each split in the tree (mean squared error was selected for this study). In RF, a random subset of the full set of predictors is selected for each node, from which a split (bootstrapping) is chosen, to ensure that influential parameters do not dominate all trees. Different hyperparameters can be used to tune the algorithm. Ntree is used to control the number of trees in the forest, mtry the number of variables available for each split and node size controls the depth of each tree. Ntree and node size were kept constant through all runs of the model, the choice of these was based on

values found in the literature [42]. The number of trees was set at 200 for the tuning phase of the model and at 1000 for training the final model, while the node size was kept at 5 for both tuning and training phase. According to Probst *et al.* (2019)[42], the performance of the model is not significantly affected by the number of trees beyond 100, the stability of the variability, however, is improved by an increase in the number of trees. For this reason 1000 trees were selected for training the final model.

The value for mtry has more influence on the performance of the model and was therefore the only tuned hyperparameter [42]. Based on early tuning attempts and the number of available predictors a range of values between 35 and 3 was selected, with most values between 25 and 14. The Out-Of-Bag Root Mean Squared Error (OOB RMSE) metric was used to choose the optimal value for mtry. The OOB RMSE is a kind of internal cross-validation where for each tree certain rows are used for training and the unused trees are used for validation (out-of-bag samples). The average of all the out-of-bag predictions can then be used to calculate a OOB RMSE.

3.2 Model setup

3.2.1 Variable selection

Different model configurations were used, including the method used by Magni *et al.* (2023)[10] as a benchmark, to investigate the benefit of using satellite products in a hybrid global streamflow modeling framework. Before choosing the different model configurations, the parameters were divided in time varying and static (describing catchment attributes) variables. Time varying variables: meteorological inputs, hydrological state variables, uncalibrated PCR-GLOBWB streamflow prediction and satellite products. Static variables: topography, river channel characteristics, soil and Groundwater properties, land use characteristics and climatic indices. More information on all the variables extracted from PCR-GLOBWB can be found in Sutanudjaja *et al.* (2018)[28] and a complete list is present in Appendix B.

Table 1 shows the different global model configurations that were used. The *pcr* configuration is the benchmark configuration, containing all the predictors obtained from PCR-GLOBWB. The other four global model configurations contain a combination of satellite products, meteorological input

and PCR-GLOBWB predictors. The *pcrSat* is used to evaluate the effect of replacing state variables related by the selected satellite products in the benchmark configuration. *PcrSatAdd* assesses the effect of adding the satellite products to the benchmark configuration.

The *satMeteo* configuration is employed to investigate what the performance of the RF model is, without input from PCR-GLOBWB. If this configuration yields good results, it would mean a significant reduction in computation time, because the run of PCR-GLOBWB is not necessary and the RF model itself is much faster. The *satMeteoStatic* configuration is subsequently used to test the change in performance after adding catchment attributes to the *satMeteo* configuration. These attributes are obtained from PCR-GLOBWB and thus would need this model to be run in advance. The attributes could, however be obtained from an external source, such as Kratzert *et al.* (2023) [43].

Table 1: Predictors used in the different global model configurations. The * indicates that the model configuration excluded: snowCoverSI, storUppTotal, storGroundwater, storLowTotal. The streamflow prediction is obtained from an uncalibrated PCR-GLOBWB.

	pcr	*pcrSat*	*pcrSat-Add*	*Sat-Meteo*	*SatMeteo-Static*
Meteorological input	X	X	X	X	X
Hydrological state variables	X	X*	X		
Streamflow prediction	X	X	X		
Satellite predictors		X	X	X	X
Static predictors	X	X	X		X

Following the global runs, the local runs were used to analyse the effectiveness of lagged variables and the effect on performance of training a model on a specific region. Canada and Australia were selected, because PCR-GLOBWB under-performed in these regions, while the United States contained a mix of stations with both good and bad performance. A region with both good and bad performance is of interest, because the hypothesized increase in performance might depend on performance in PCR-GLOBWB. A RF model was trained for each region separately.

Table 2 shows the local model configurations that were used. Lagged values of LWE and SCF were added for the *pcrSatLag* configurations, while the *satMeteoLag* configurations also included lagged values for the meteorological inputs. The SM values contained too many missing values to be of use in a lagged simulation, as rows with missing data would be multiplied by the selected lag number.

Table 2: Predictors used in different local model configurations. The * indicates that the model configuration excluded: snowCoverSWE, storUppTotal, storGroundwater, storLowTotal. Lag is added to LWE and SCF for *pcrSatLag*, *satMeteoLag* lag for the meteorological input variables. The streamflow prediction is obtained from an uncalibrated PCR-GLOBWB.

	pcr	*pcr-Sat*	*pcr-Sat-Lag4*	*pcr-Sat-Lag12*	*Sat-Meteo*	*Sat Meteo-Lag4*	*Sat-Meteo-Lag12*
Meteorological inputs	X	X	X	X	X	X	X
Hydrological state variables	X	X*	X*	X*			
Streamflow prediction	X	X	X	X			
Satellite predictors		X	X	X	X	X	X
Static predictors	X	X	X	X			
Lag 4 months			X	X		X	
Lag 12 month				X			X

3.2.2 Training and evaluation

The *Ranger* package [44] was used to implement RF in R, which is capable of fast implementation with high-dimensional data. The RF model requires three hyperparameters (ntree, node size mtry) to be tuned. To reduce computation time ntree and node size were kept constant, while only tuning mtry (see Section 3.1).

Five random subsamples were created for training and testing. The train-test split was achieved with a location-based approach, with about 70% of the stations used for training and 30% for testing. Each model configuration in the same region (global, Australia, Canada or United States) contained the same stations for training and testing. All data for training and testing was aggregated in a single table to train the RF. An additional train-test split condition was applied to ensure that about two-thirds of the observations was present in the training set. This additional condition is needed due to the high level of missing observations for some stations. After training the RF model, new streamflow predictions are created for each station in the test-set.

Following model training, the contribution of variables to a final prediction can be determined by assessing their variable importance [45]. This study used the 'impurity' measure, based on the Gini criterion, which calculates the average decrease in Gini node-impurity over all trees per variable.

The performance is assessed individually for each station, using a station-by-station approach. The Kling-Gupta efficiency (KGE) is selected for evaluating the performance of the model [46]. KGE is calculated by combining three different components (see Equation 2): the linear correlation between observation and prediction (r), the variability in the flow (σ indicating the standard deviation) and a bias term (μ indicating the mean). KGE values range from $-\infty$ to 1, with scores above -0.41 indicating a good KGE value [47].

$$KGE = 1 - \sqrt{(r-1)^2 + (\frac{\sigma_{pred}}{\sigma_{obs}} - 1)^2 + (\frac{\mu_{pred}}{\mu_{obs}} - 1)^2} \tag{2}$$

4. Results

The results are organized in two sections. First results for the global model configurations are shown (see section 4.1, followed by the local results (see section 4.2.

4.1 Global runs

4.1.1 Variable importance

Figure 5 displays the variable importance for the five global configurations averaged over the five subsamples. The figure only shows the 20 most important variables for each configuration. To ensure that even small changes are discernible, all variables have been scaled to their square root.

The *pcr*, *pcrSat* and *pcrSatAdd* configurations present quite similar results for most of the variables. It is, however, clear that the satellite products LWE and SM are important and that SM is the most important satellite feature to be added to the model. Similar to the results found by Magni *et al.* (2023)[10], the flow depth from PCR-GLOBWB is the most important predictor for all configurations that include this predictor.

The *satMeteo* configuration shows that the satellite products are more important compared to the meteorological input predictors. This changes with the addition of the static variables, in *satMeteoStatic*. The precipitation is now more important than SCF. The error bars, in Figure 5, show the relative uncertainty of each predictor, which remains roughly constant for all configurations.

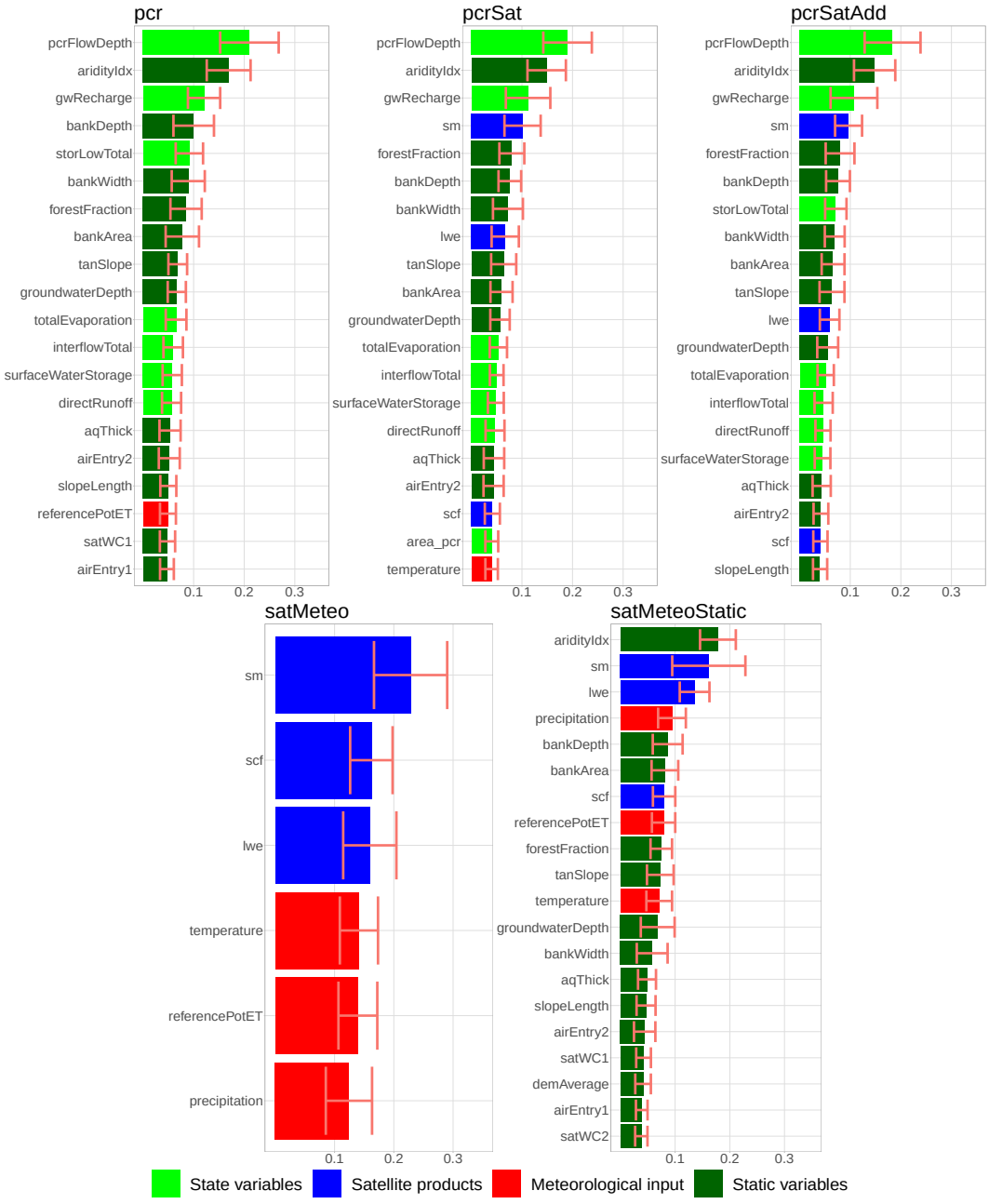

Figure 5: Square rooted mean decrease in impurity values of the top twenty variables. The figure includes all five global RF configurations averaged over the five subsamples. The different type of variables are indicated by color.

4.1.2 Tuning

Figure 6 shows the results of tuning the mtry hyperparamter, which was performed for each subsample and each configuration. For all configurations the ntry was kept constant at 200 and the node size at 5. The number of tuned values depends on the number of predictors used in the configuration. The configurations *pcr*, *pcrSat* and *pcrSatAdd* display small fluctuations around the minimum OOB RMSE, with quite large differences between each of the subsamples. The *satMeteoStatic* configuration shows less fluctuation around the minimum OOB RMSE and contains far smaller differences between the configurations. The *satMeteo* configuration is only tuned for two values of mtry and shows very little difference between the two values.

All configurations in figure 6, except for *satMeteo*, show a rapid decrease followed by a slight increase. The increase is caused by overfitting on the training data. The lowest values for each configuration and subsample were selected and used for training the final model (see Table 3).

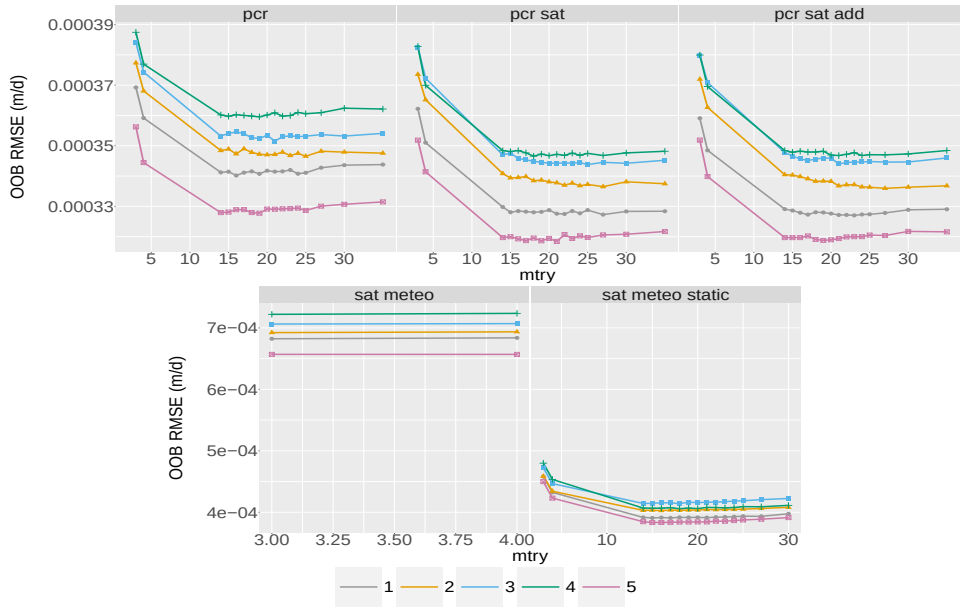

Figure 6: RF tuning of mtry hyperparameter. Each panel shows all subsamples and a single configuration. The lines indicate a OOB RMSE score and the dots indicate the tuned mtry values. Using a fixed ntree of 200 and node size of 5.

Table 3: Optimal mtry values, obtained from tuning the following values: 35,30,27,25,24,23,22,21,20,19,18,17,16,15,14,4,3. *satMeteo* and *satMeteoStatic* are limited by the number of available variables. Constant ntree of 200.

Subsample / **Configuration**	*1*	*2*	*3*	*4*	*5*
pcr	16	25	21	19	19
pcrSat	27	27	25	18	21
pcrSatAdd	23	27	21	21	19
satMeteo	3	3	3	3	4
SatMeteo-Static	15	16	18	18	15

4.1.3 Performance

Boxplots for all global model configurations of KGE scores are visualized in Figure 7. Each row in Figure 7 represents a subsample or the cumulative of all subsamples (row indicated with C). Within each row, each boxplot displays the performance of different configurations and the uncalibrated PCR-GLOBWB prediction.

It is clear in Figure 7 that the performance of the different model configurations is not dependent on the selected subsample, which indicates that the model is not overfitting on a specific region or set of stations. The configurations *pcr*, *pcrSat*, *pcrSatAdd* and *satMeteoStatic* all show a significant improvement in performance compared to the uncalibrated PCR-GLOBWB (red boxplot). The four configurations also show very little differences between them, for all rows. There are only slight variations visible in the outliers for some of the subsamples, which are smoothed out in the cumulative.

The *satMeteo* configuration also shows significant improvement compared to the uncalibrated PCR-GLOBWB configuration. This configuration, however, shows a larger spread and an overall lower performance compared to other four configurations.

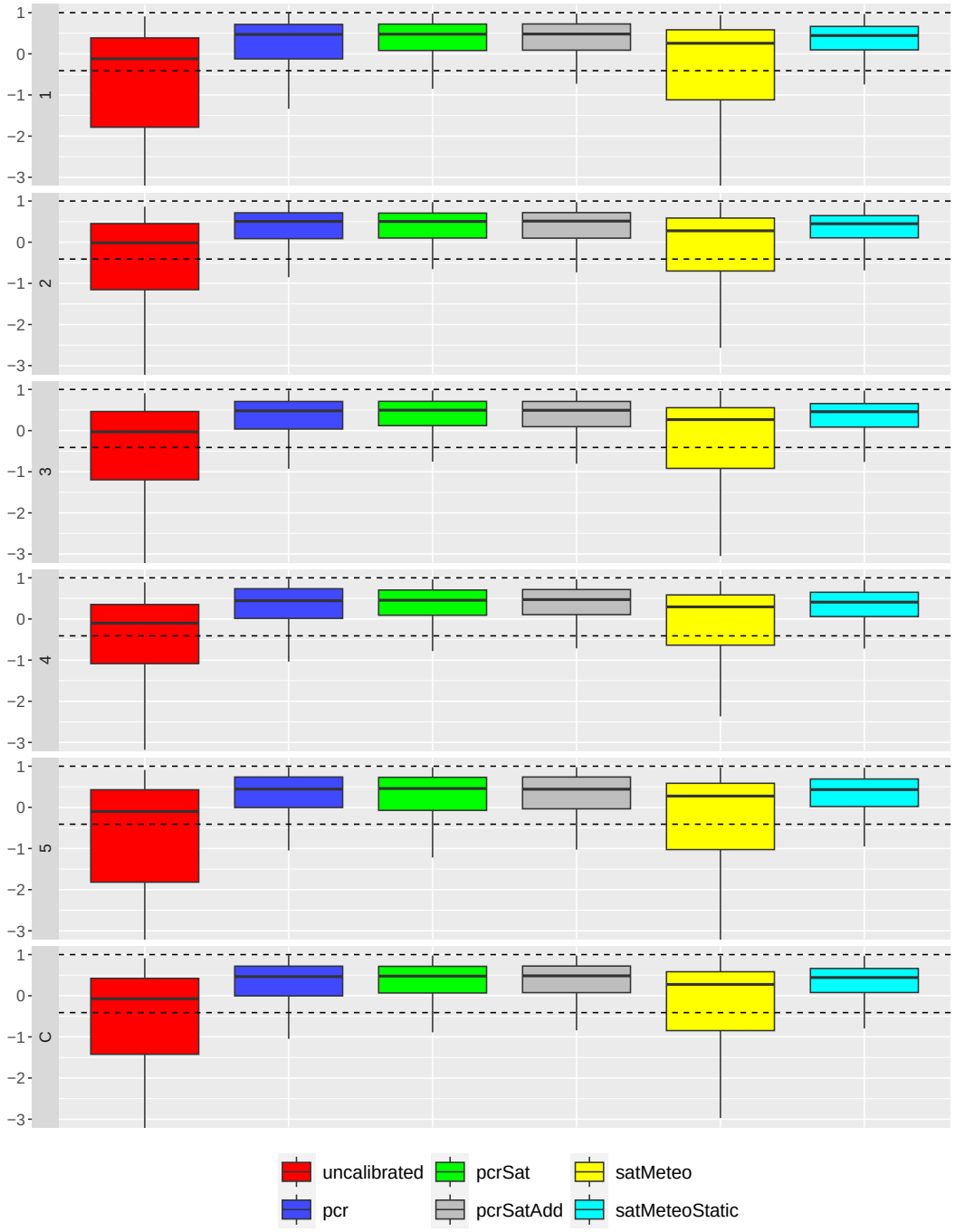

Figure 7: Boxplots of KGE for all five subsamples and the uncalibrated PCR-GLOBWB. The two dotted lines indicate an ideal KGE value of 1 and the minimum "good" value of -0.41. The values are limited to -3 on the y-axis

Figure 8 presents the cumulative distribution of KGE score averaged over the five subsamples. Each line in the figure indicates a different configuration or the uncalibrated PCR-GLOBWB.

The cumulative distribution largely shows a similar pattern, as seen in Figure 7, with all five performing significantly better compared to the uncalibrated PCR-GLOBWB and *satMeteo* showing lower performance compared to the other four configurations. The only difference between *pcr*, *pcrSat*, *pcrSatAdd* and *satMeteoStatic* is that below 0 *satMeteoStatic* performs slightly better compared to the other three, while the performance is very slightly worse above 0. This results in median values of 0.478, 0.486, 0.488 and 0.431 for *pcr*, *pcrSat*, *pcrSatAdd* and *satMeteoStatic* respectively.

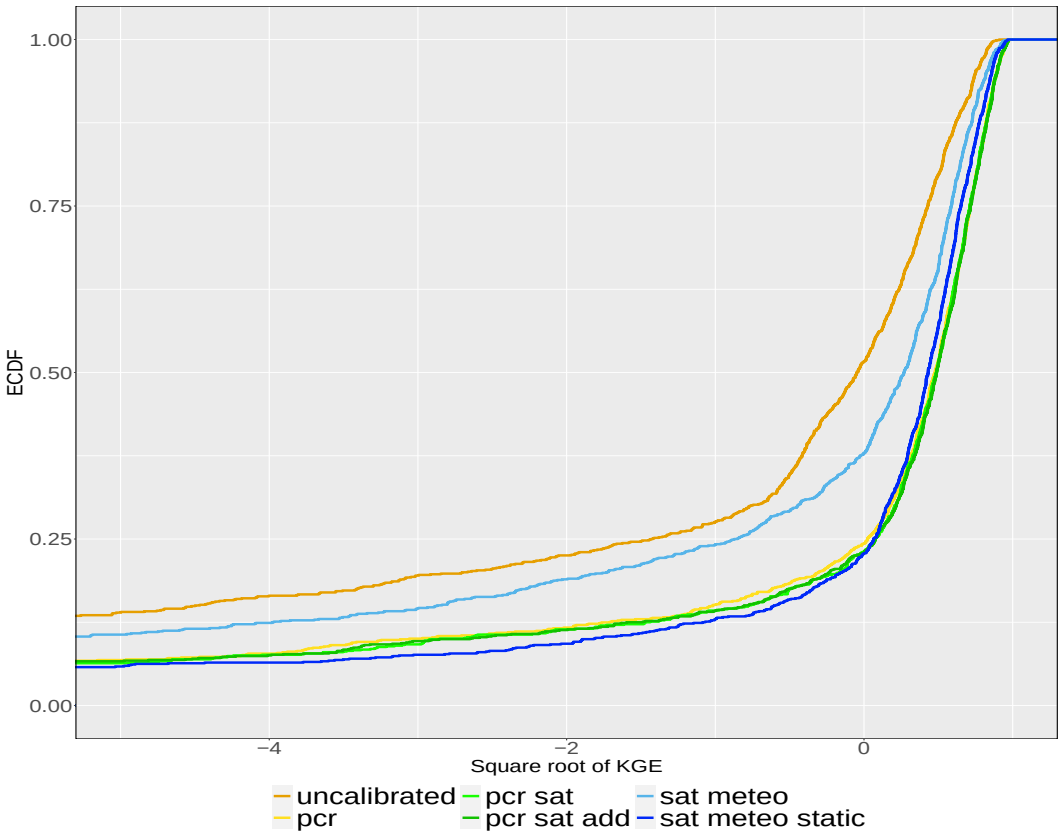

Figure 8: Cumulative distribution functions of KGE for the five configurations and the uncalibrated PCR-GLOBWB. The KGE results were averaged over the five subsamples and the lower KGE scores are not shown and the x-axis limited to -5.

Table 4 shows that even though the cumulative distribution for KGE is very similar for some of the configurations, there are still large differences for individual stations or catchments. Row 1 and 2 show stations (in Spain and Ghana), where the *satMeteo* and *satMeteoStatic* configurations significantly outperform the the other configurations. The rows three and four show two stations (in the United States) for the opposite case where *pcr*, *pcrSat* and *pcrSatAdd* outperform the other two configurations. For both cases the uncalibrated PCR-GLOBWB has KGE values below -0.41.

Table 4: KGE values for individual stations (columns indicate GRDC stations) averaged over the five subsamples for each configuration (rows). GH stand for Ghana, ES for Spain and US for United States.

Config. / **Station**	*Uncalibrated PCR-GLOBWB*	*pcr*	*pcrSat*	*pcrSat-Add*	*Sat-Meteo*	*Sat-Meteo-Static*
1531650 (GH)	-5.87	-0.53	-0.66	-0.48	0.49	0.60
6226600 (ES)	-12.01	-0.72	-0.76	-0.63	0.02	0.44
4122600 (US)	-0.81	0.81	0.85	0.90	-0.69	-0.71
4122701 (US)	-2.30	0.58	0.71	0.69	-0.64	-0.41

4.2 Local runs

This section presents the results obtained from training the models on the stations in Australia, Canada, and the United States. The variable importance is different for the locally trained models compared to the global model (See Figure D.1 for Australia, Figure D.2 for Canada and Figure D.3 for the United States). Similar to the global results, is that LWE and SM are still high among the 20 most important predictors for Australia and the United States. Canada, however, shows a difference in the importance of the satellite products, where the SCF is the most important satellite product added to the configurations *pcrSat* and *pcrSatAdd*. Further investigation of the differences in variable importance fall outside the scope of this study.

It was found that there was no overfitting to specific stations for Canada and the United States (see Figures D.5 and D.6). Australia, however, showed that there was overfitting, because there were significant differences between the results for each subsamples (see figures D.4).

Figure 9 shows the cumulative distribution of the KGE scores, in which the results for the 12 month lag are not included, because this configuration did not show large differences with the 4 month lag (see Figure C.1). Similarly to the global runs (see Figure 8) all configurations for Canada and Australia show considerable improvement compared to the uncalibrated PCR-GLOBWB run. The cumulative distribution of Australia, however, is characterized by a very irregular pattern. It is difficult to visually find a clear pattern between the different configurations. However the median values indicate that *pcrSat* has the best performance (see Table 5).

Table 5: Median KGE values averaged over the five subsamples. The rows indicate the countries and the columns the different configurations

	Uncalibrated PCR-GLOBWB	*pcr*	*pcrSat*	*pcrSat-Lag4*	*Sat-Meteo*	*Sat-Meteo-Lag4*
Australia	-0.10	-0.10	0.15	-0.01	-0.12	-0.1
Canada	0.15	0.53	0.54	0.51	0.34	0.40
United States	-0.14	0.52	0.51	0.51	0.27	0.35

Upon further visual inspection of Figure 9 it was found that for Canada and the United States there is very little difference between the *pcr*, *pcrSat* and *pcrSatLag4* configurations. For both countries the added lag slightly improves the *pcrMeteo* configurations, which is also visible in the median values for these configurations (see Table 5). For Australia it is unclear from the figure if there is a difference with the addition of lag.

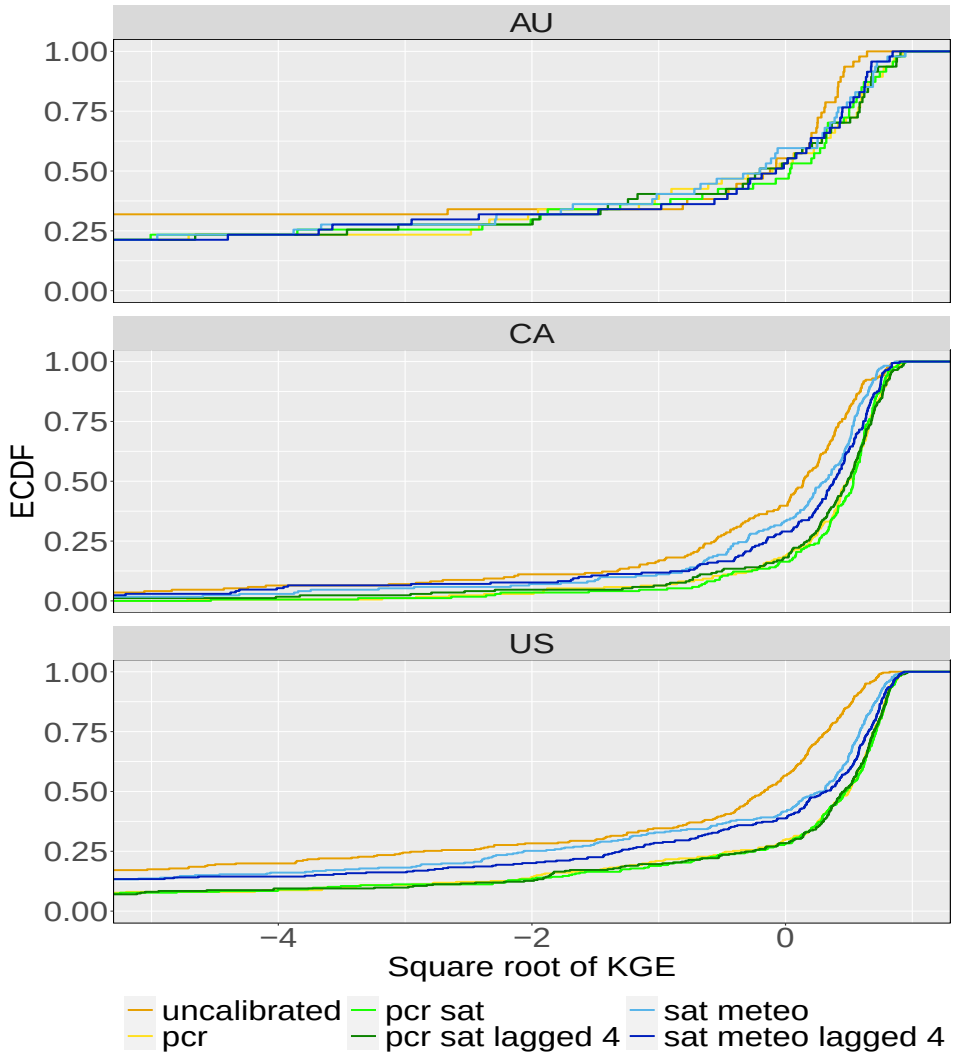

Figure 9: Cumulative distribution function of KGE for five local configurations and the uncalibrated PCR-GLOBWB. The results were averaged over the five subsamples. AU indicates Australia, CA Canada and US United States.

5. Discussion

The results in this study proved that the performance of a hybrid global streamflow framework, with RF and PCR-GLOBWB, is not improved for global runs or local runs in Australia, Canada and the United States. On the other hand, it has been proven that the partial or complete removal of variables from PCR-GLOBWB can still retain high performance in the RF post-processor. Both the configuration only satellite products and meteorological input and the configuration with satellite products, meteorological and catchment attributes show significant improvement compared to an uncalibrated PCR-GLOBWB. The configuration including catchment attributes even showed comparable performance to configurations with input of state variables from PCR-GLOBWB. There are many studies in a review by Jiang and Wang (2019)[12] stating either the benefit or drawbacks of using remote sensing in streamflow predictions. This study however showed that there are clear advantages to using remote sensing data in the proposed hybrid framework. Before the configuration with satellite products, meteorological input and catchment attributes can used to improve upon the method proposed by Magni *et al.* (2023)[10], catchment attributes from an external source need to be used, such as those from Kratzert *et al.* (2023)[43].

The added lagged variables did not significantly increase the overall performance of the RF post-processor. We did, however find a slight difference between the *satMeteo* and *satMeteoLag4* for Canada and the United States. Additional lag of 12 months did not change the performance and significantly increases the number of missing values making the model less reliable. Other studies have found that the addition of lagged variables in hydrological models can improve performance [48]. The differnce with findings in this study could be explained by the fact that RF is not an ideal statistical learning algorithm to include lag, since it only adds a variable to each row as an extra predictor, instead of looking for a pattern in the time series. Different algorithms like the Long-short-term-memory (LSTM) algorithm are better suited to test the effect of adding lag [49].

To further analyse the differences in performance for configurations with

satellite products, local effects were investigated. These showed that the models trained on different regions contain considerable differences, from the global simulations. Part of these difference can be explained by a lack of station values and therefore might not be statistically significant. This can be seen in the model for Australia that contained a limited number of stations, which could be a reason for the difference in performance.

5.1 Limitations and recommendations

The selection of predictors to be removed from the set of PCR-GLOBWB predictors was based on their relation with the satellite products, but has not been extensively tested. Further combinations of predictors should be tested to find what the effect is of removing these specific predictors. Additionally, there are many more available remote sensing products that can improve streamflow predictors [12]. Two possible products are are evapotranspiration data from the Global Land Evaporation Amsterdam Model [50] and precipitation from the Soil Moisture to Rain dataset [51].

For the research into local effects, three countries with low or mixed performance in the PCR-GLOBWB were selected. But the performance of hydrological models is not dependent on national borders. This selection of countries is therefore not suitable for investigating the performance of the model in different regions. Further research, should focus on clustering approaches of relevant predictors to find regions with similar hydrological characteristics.

Finally, the effect of missing data has not been considered in this study. There are many months with missing data in the GRDC observations and SM satellite data. The mechanism of missing data is an important aspect of the effect of missing data [52], but has not been investigated. A solution might be the application of a RF-based mixed effects model, which are often used studies with repeated measurements on the same object. This type of model is capable of dealing with time-series analysis, where there are uneven records [53]. Additionally, to reduce the number of missing values in the satellite observations, a method was used where missing values were changed to zero. This approach reduces the overall averaged values and more investigation into the effect of this method is needed.

6. Conclusion

This study showed that adding LWE, SCF and SM satellite data to a hybrid framework using Random Forest and PCR-GLOBWB, does not improve streamflow predictions on a global scale. However, satellite data, meteorological data and catchment attributes can be used to create an accurate model, with a relative small number of predictors. Additionally, a model with only six predictors (three satellite and three meteorological) can already increase performance. This means that a small model can be used for quick streamflow predictions.

It has also been proven that a change in performance is dependent on the location of the upstream basin and that the addition of lag does not significantly improve predictions, when using a RF algorithm. This limitation is likely caused by, how the RF algorithm gains additional information from the lagged predictors. Therefore the effectiveness of lag cannot be truly judged, based on the method used in this study

Future research may focus on: 1) increasing the number of predictor variables extracted from satellite products to the hybrid framework; 2) cluster-analysing stations where the addition of satellite products improves the post-processor compared to the method proposed by Magni *et al.* (2023) [10] ; 3) applying different SL algorithms that are better equipped to deal with lagged variables; 4) using mixed effects modelling.

A. Appendix

All the codes used in this research can be found at: https://github.com/niekcde/PCR-GLOBWB-satellite-RF-reanalysis.

B. Appendix

Table B.1: All variables extracted from PCR-GLOBWB and used in the RF model. Further description of the predictors can be found in

Variable	Unit	Description	Type of predictor
pcrFlowDepth	m/day	PCR-GLOBWB daily streamflow predictions	Varying
precipitation	m/day	Meteorological Input	Varying
temperature	°C	Meteorological input	Varying
referencePotET	m/day	Meteorological input	Varying
baseflow	m/day	State variable	Varying
desalinationAbstraction	m/day	State variable	Varying
directRunoff	m/day	State variable	Varying
domesticWaterWithdrawal	m/day	State variable	Varying
fossilGroundwaterAbstraction	m/day	State variable	Varying
gwRecharge	m/day	State variable	Varying
industryWaterWithdrawal	m/day	State variable	Varying
interflowTotal	m/day	State variable	Varying
irrigationWaterWithdrawal	m/day	State variable	Varying
livestockWaterWithdrawal	m/day	State variable	Varying
nonIrrWaterConsumption	m/day	State variable	Varying
snowCoverSWE	m	State variable	Varying
snowFreeWater	m	State variable	Varying
storGroundwater	m	State variable	Varying
storLowTotal	m	State variable	Varying
storUppTotal	m	State variable	Varying
surfaceWaterAbstraction	m/day	State variable	Varying
surfaceWaterInf	m/day	State variable	Varying
surfaceWaterStorage	m	State variable	Varying

Table B.1 continued from previous page

totalGroundwaterAbstraction	m/day	State variable	Varying
totalEvaporation	m/day	State variable	Varying
airEntry1		Catchment attributes	Static
airEntry2		Catchment attributes	Static
aqThick		Catchment attributes	Static
aridityIdx		Catchment attributes	Static
bankArea	m2	Catchment attributes	Static
bankDepth	m	Catchment attributes	Static
bankWidth	m	Catchment attributes	Static
demAverage	m	Catchment attributes	Static
forestFraction		Catchment attributes	Static
groundwaterDepth		Catchment attributes	Static
KSat1		Catchment attributes	Static
KSat2		Catchment attributes	Static
kSatAquifer		Catchment attributes	Static
recessionCoeff		Catchment attributes	Static
resWC1		Catchment attributes	Static
resWC2		Catchment attributes	Static
satWC1		Catchment attributes	Static
satWC2		Catchment attributes	Static
slopeLength		Catchment attributes	Static
specificYield		Catchment attributes	Static
Storage1		Catchment attributes	Static
storage2		Catchment attributes	Static
storDepth1		Catchment attributes	Static
storDepth2		Catchment attributes	Static
tanSlope		Catchment attributes	Static
poreSize1		Catchment attributes	Static
poreSize2		Catchment attributes	Static
percolationImp		Catchment attributes	Static
catchment area	m2	Catchment attributes	Static

C. Appendix

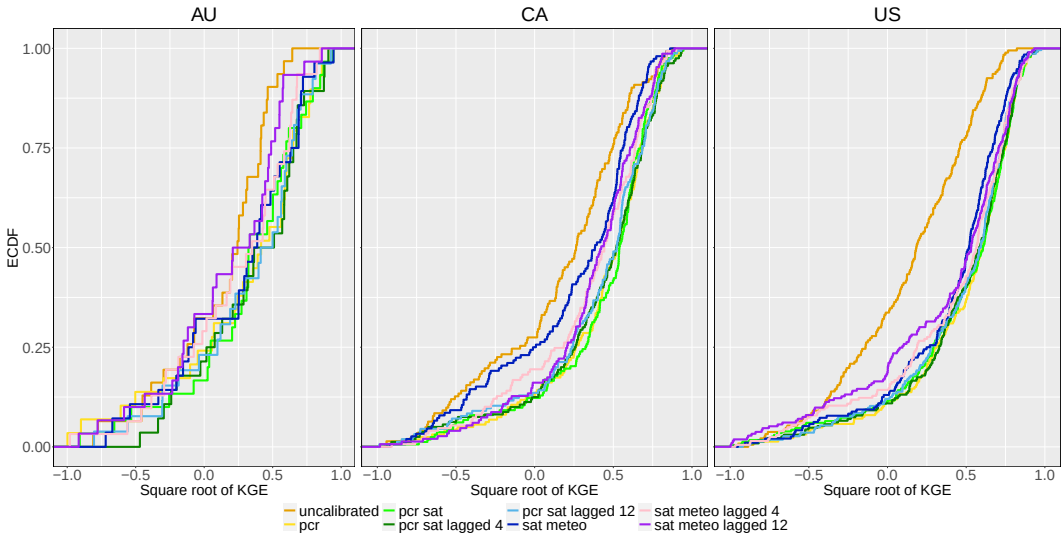

Figure C.1: Cumulative distribution function of KGE for the five local model configurations shown in the results section, the uncalibrated PCR-GLOBWB and the 12 month lag. The results were averaged over the five subsamples.

D. Appendix

D.1 Variable importance local configurations

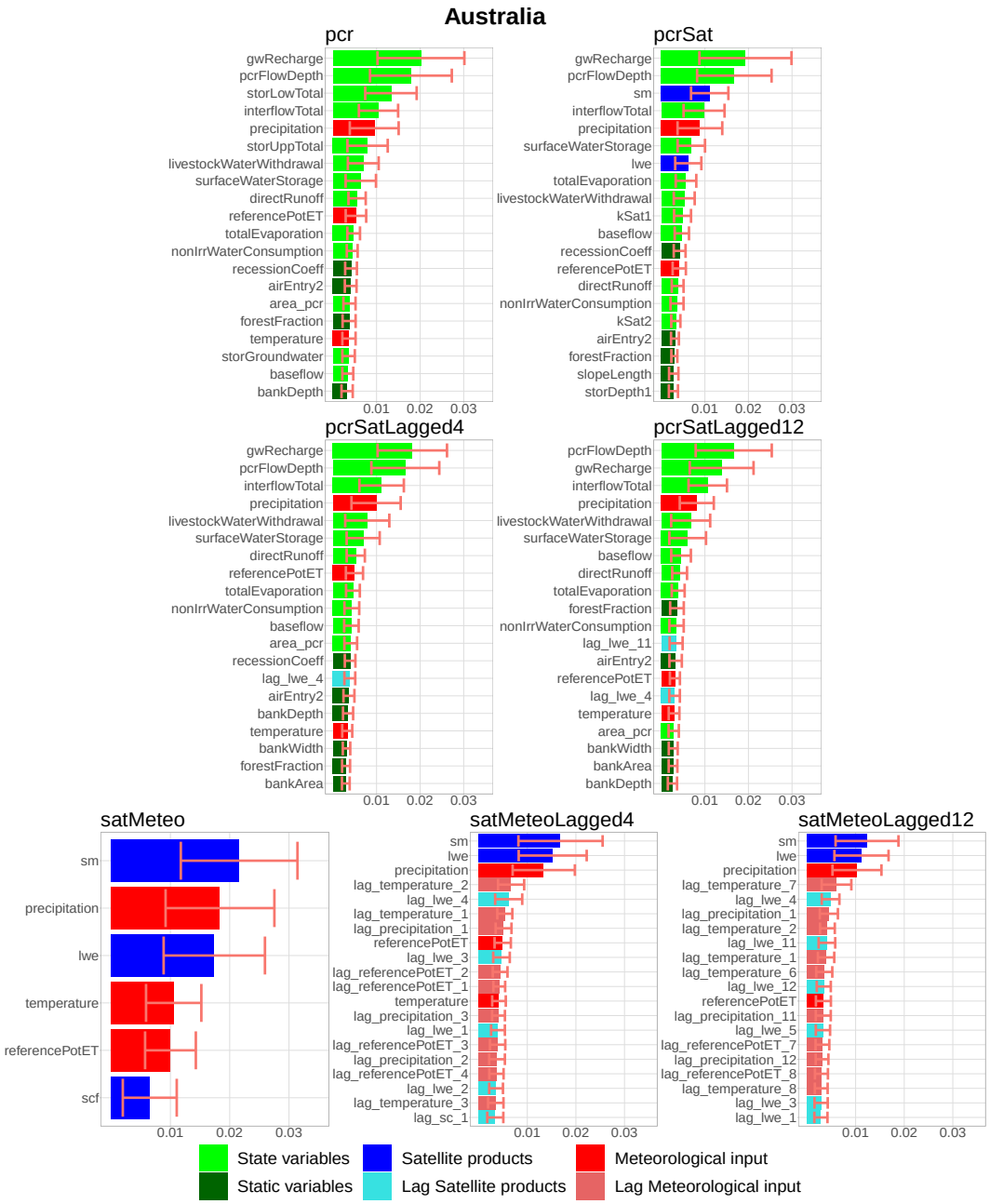

Figure D.1: Variable importance for Australia. Square rooted mean decrease in impurity values of the top twenty variables. The figure includes all seven local RF configurations averaged over the five subsamples. The different type of variables are indicated by color.

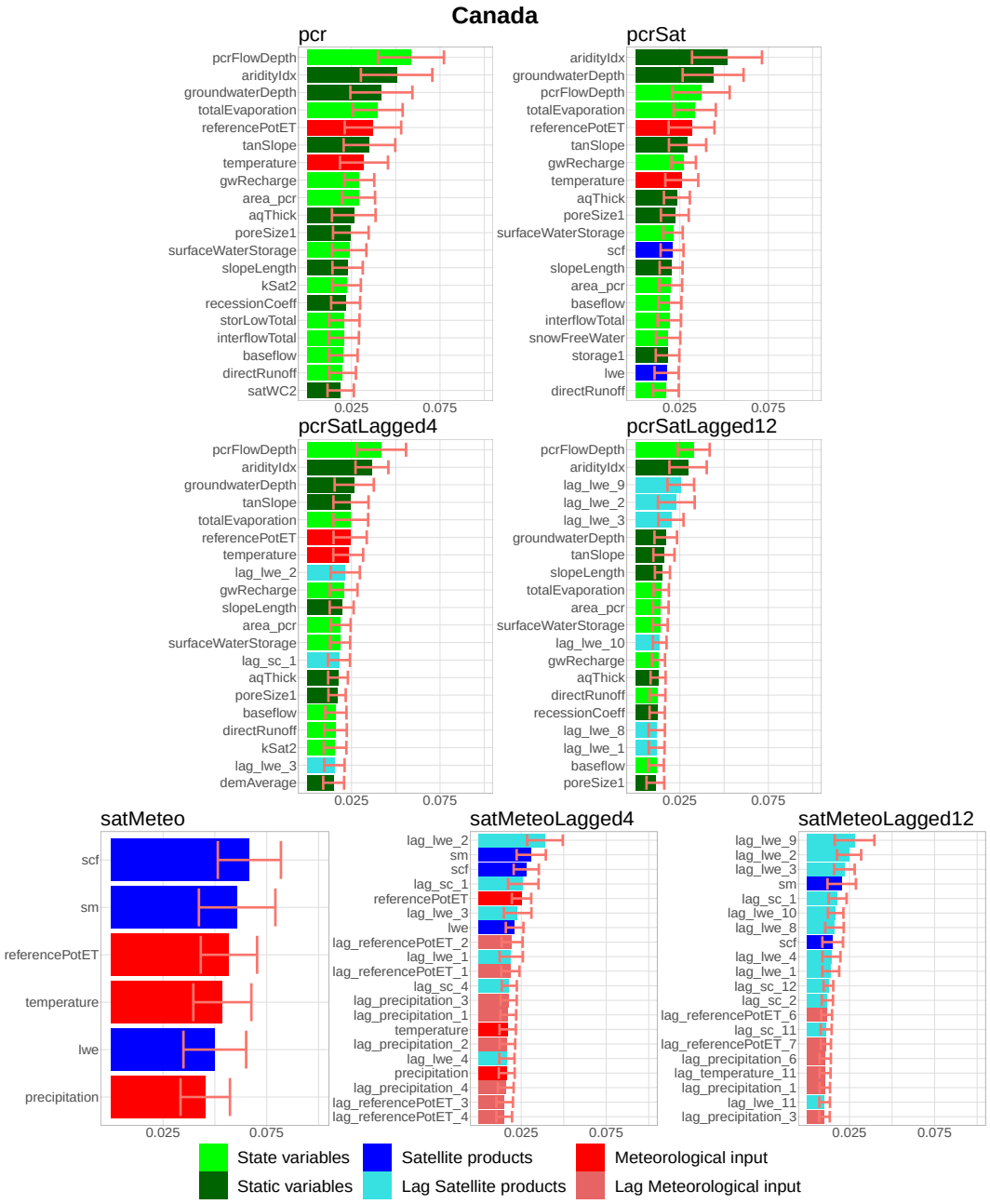

Figure D.2: Variable importance for Canada. Square rooted mean decrease in impurity values of the top twenty variables. The figure includes all seven local RF configurations averaged over the five subsamples. The different type of variables are indicated by color.

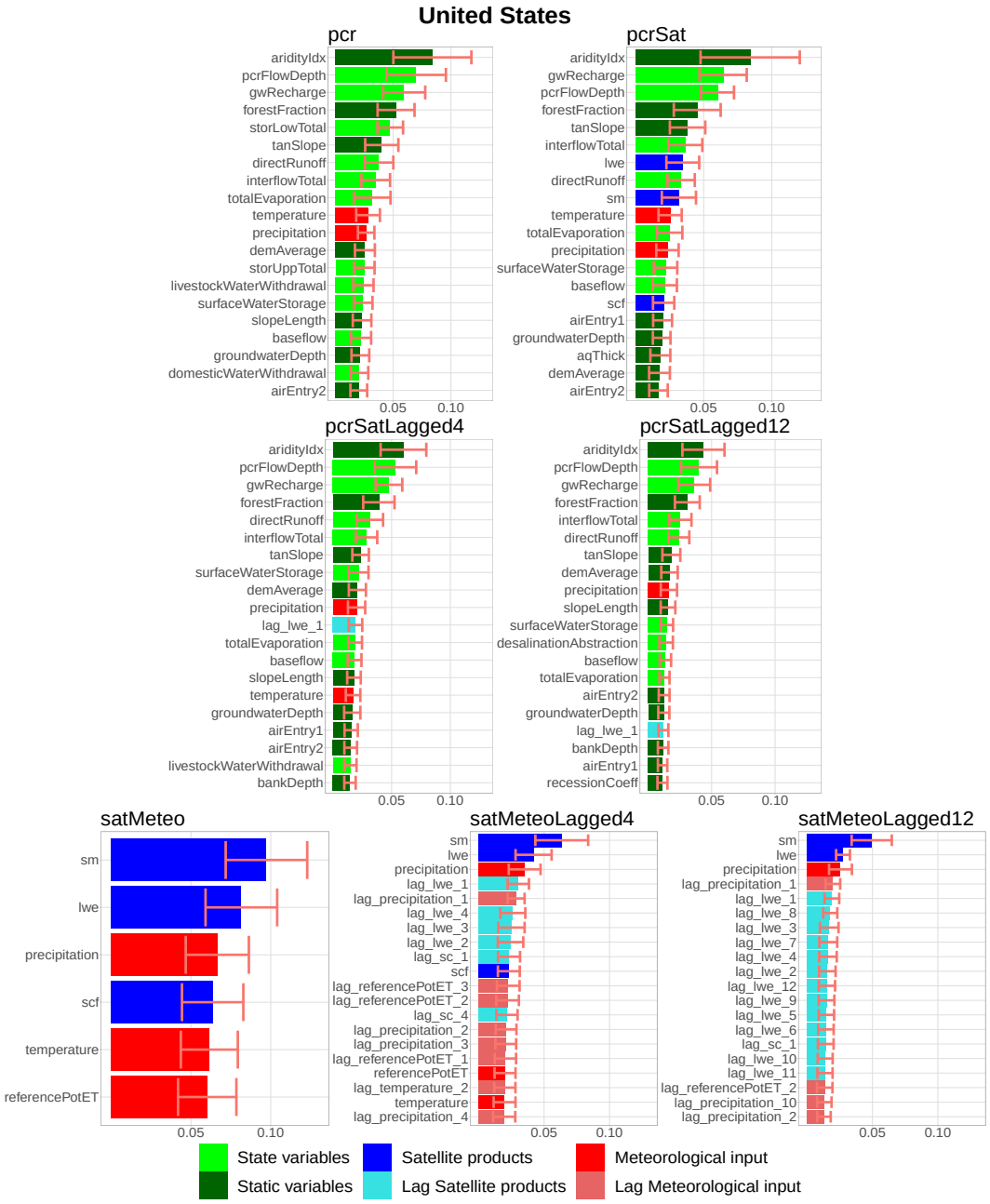

Figure D.3: Variable importance for the United States. Square rooted mean decrease in impurity values of the top twenty variables. The figure includes all seven local RF configurations averaged over the five subsamples. The different type of variables are indicated by color.

D.2 Boxplots local configurations

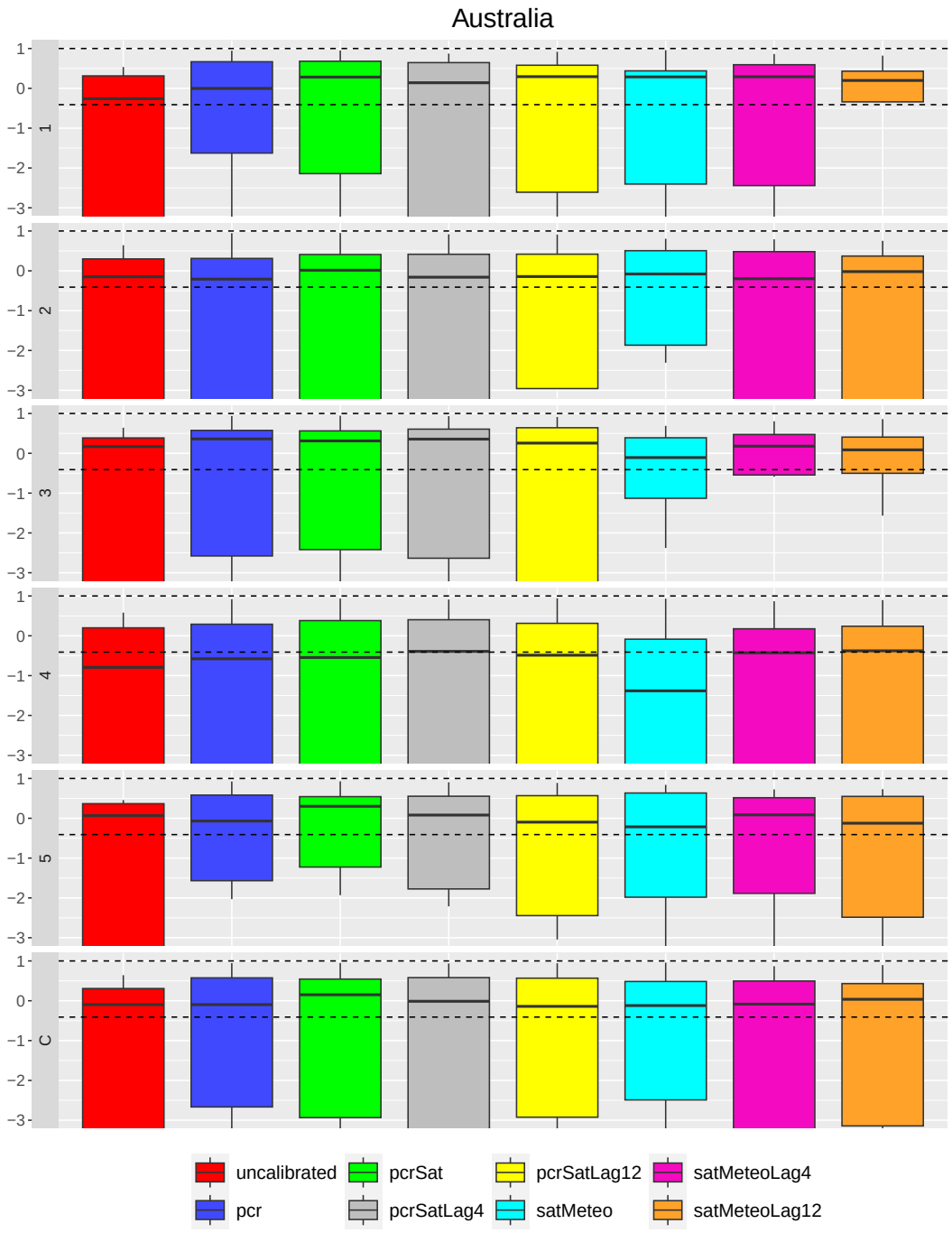

Figure D.4: Boxplots for the region Australia of KGE for all five subsamples and the uncalibrated PCR-GLOBWB. The two dotted lines indicate an ideal KGE value of 1 and the minimum "good" of -0.41.

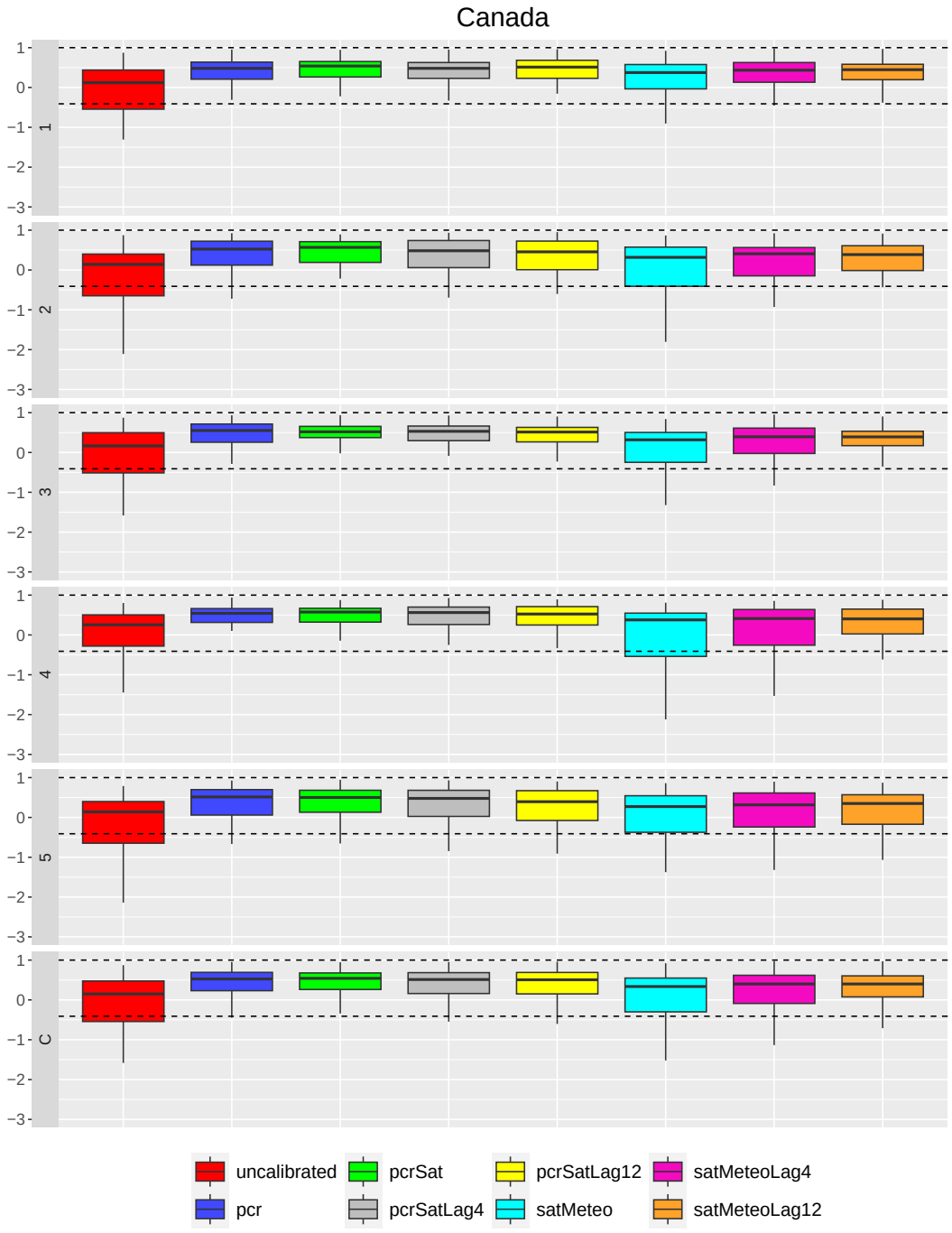

Figure D.5: Boxplots for the region Canada of KGE for all five subsamples and the uncalibrated PCR-GLOBWB. The two dotted lines indicate an ideal KGE value of 1 and the minimum "good" of -0.41.

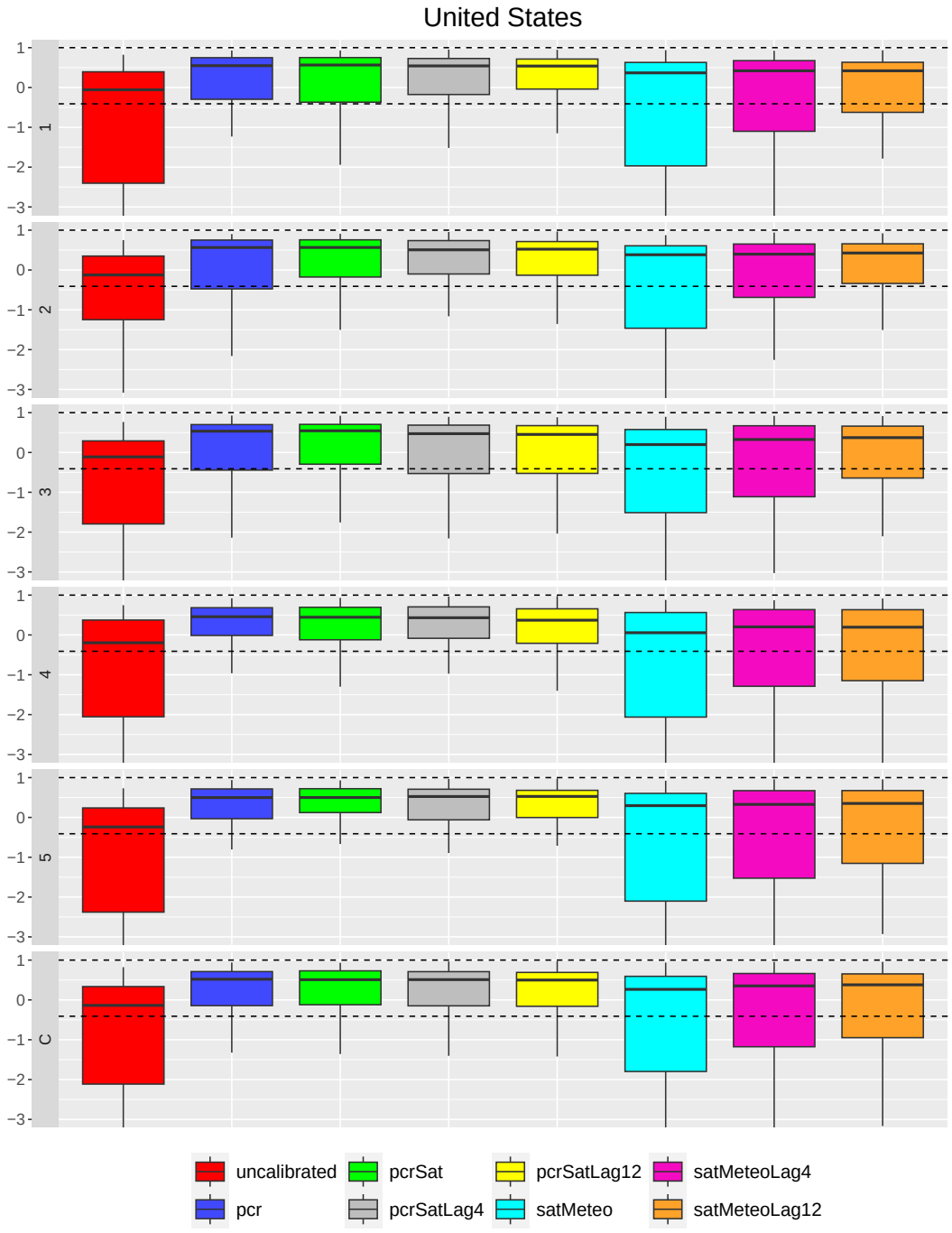

Figure D.6: Boxplots for the region United States of KGE for all five subsamples and the uncalibrated PCR-GLOBWB. The two dotted lines indicate an ideal KGE value of 1 and the minimum "good" of -0.41.

Bibliography

[1] Z. W. Kundzewicz, "Climate change impacts on the hydrological cycle," *Ecohydrology &amp Hydrobiology*, vol. 8, no. 2-4, pp. 195–203, Jan. 2008. DOI: 10.2478/v10104-009-0015-y. [Online]. Available: https://doi.org/10.2478/v10104-009-0015-y.

[2] A. Patil and R. Ramsankaran, "Improving streamflow simulations and forecasting performance of SWAT model by assimilating remotely sensed soil moisture observations," *Journal of Hydrology*, vol. 555, pp. 683–696, Dec. 2017. DOI: 10.1016/j.jhydrol.2017.10.058. [Online]. Available: https://doi.org/10.1016/j.jhydrol.2017.10.058.

[3] P. Laiolo, S. Gabellani, L. Campo, *et al.*, "Impact of different satellite soil moisture products on the predictions of a continuous distributed hydrological model," *International Journal of Applied Earth Observation and Geoinformation*, vol. 48, pp. 131–145, Jun. 2016. DOI: 10.1016/j.jag.2015.06.002. [Online]. Available: https://doi.org/10.1016/j.jag.2015.06.002.

[4] Y. Liu and H. V. Gupta, "Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework," *Water Resources Research*, vol. 43, no. 7, Jul. 2007. DOI: 10.1029/2006wr005756. [Online]. Available: https://doi.org/10.1029/2006wr005756.

[5] H. V. Gupta, S. Sorooshian, and P. O. Yapo, "Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information," *Water Resources Research*, vol. 34, no. 4, pp. 751–763, Apr. 1998. DOI: 10.1029/97wr03495. [Online]. Available: https://doi.org/10.1029/97wr03495.

[6] Q. Duan, S. Sorooshian, and V. K. Gupta, "Optimal use of the SCE-UA global optimization method for calibrating watershed models," *Journal of Hydrology*, vol. 158, no. 3-4, pp. 265–284, Jun. 1994. DOI: 10.1016/0022-1694(94)90057-4. [Online]. Available: https://doi.org/10.1016/0022-1694(94)90057-4.

[7] B. A. Tolson and C. A. Shoemaker, "Dynamically dimensioned search algorithm for computationally efficient watershed model calibration," *Water Resources Research*, vol. 43, no. 1, Jan. 2007. DOI: 10.1029/2005wr004723. [Online]. Available: https://doi.org/10.1029/2005wr004723.

[8] Y. Shen, J. Ruijsch, M. Lu, E. H. Sutanudjaja, and D. Karssenberg, "Random forests-based error-correction of streamflow from a large-scale hydrological model: Using model state variables to estimate error terms," *Computers &amp Geosciences*, vol. 159, p. 105 019, Feb. 2022. DOI: 10.1016/j.cageo.2021.105019. [Online]. Available: https://doi.org/10.1016/j.cageo.2021.105019.

[9] B. Kraft, M. Jung, M. Körner, S. Koirala, and M. Reichstein, "Towards hybrid modeling of the global hydrological cycle," *Hydrology and Earth System Sciences*, vol. 26, no. 6, pp. 1579–1614, Mar. 2022. DOI: 10.5194/hess-26-1579-2022. [Online]. Available: https://doi.org/10.5194/hess-26-1579-2022.

[10] M. Magni, E. H. Sutanudjaja, Y. Shen, and D. Karssenberg, "Global streamflow modelling using process-informed machine learning," Feb. 2023. DOI: 10.5194/egusphere-egu23-1160. [Online]. Available: https://doi.org/10.5194/egusphere-egu23-1160.

[11] S. C. Worland, W. H. Farmer, and J. E. Kiang, "Improving predictions of hydrological low-flow indices in ungaged basins using machine learning," *Environmental Modelling &amp Software*, vol. 101, pp. 169–182, Mar. 2018. DOI: 10.1016/j.envsoft.2017.12.021. [Online]. Available: https://doi.org/10.1016/j.envsoft.2017.12.021.

[12] D. Jiang and K. Wang, "The role of satellite-based remote sensing in improving simulated streamflow: A review," *Water*, vol. 11, no. 8, p. 1615, Aug. 2019. DOI: 10.3390/w11081615. [Online]. Available: https://doi.org/10.3390/w11081615.

[13] R. Anyah, E. Forootan, J. Awange, and M. Khaki, "Understanding linkages between global climate indices and terrestrial water storage changes over africa using GRACE products," *Science of The Total Environment*, vol. 635, pp. 1405–1416, Sep. 2018. DOI: 10.1016/j.scitotenv.2018.04.159. [Online]. Available: https://doi.org/10.1016/j.scitotenv.2018.04.159.

[14] R. Garcıa-Herrera and D. Barriopedro, "Northern hemisphere snow cover and atmospheric blocking variability," *Journal of Geophysical Research*, vol. 111, no. D21, Nov. 2006. DOI: 10.1029/2005jd006975. [Online]. Available: https://doi.org/10.1029/2005jd006975.

[15] C. D. U. Carranza, M. J. van der Ploeg, and P. J. J. F. Torfs, "Using lagged dependence to identify (de)coupled surface and subsurface soil moisture values," *Hydrology and Earth System Sciences*, vol. 22, no. 4, pp. 2255–2267, Apr. 2018. DOI: 10.5194/hess-22-2255-2018. [Online]. Available: https://doi.org/10.5194/hess-22-2255-2018.

[16] A. Hammer, D. Heinemann, E. Lorenz, and B. Lückehe, "Short-term forecasting of solar radiation: A statistical approach using satellite data," *Solar Energy*, vol. 67, no. 1-3, pp. 139–150, Jul. 1999. DOI: 10.1016/s0038-092x(00)00038-4. [Online]. Available: https://doi.org/10.1016/s0038-092x(00)00038-4.

[17] D. I. F. Grimes, E. Coppola, M. Verdecchia, and G. Visconti, "A neural network approach to real-time rainfall estimation for africa using satellite data," *Journal of Hydrometeorology*, vol. 4, no. 6, pp. 1119–1133, Dec. 2003. DOI: 10.1175/1525-7541(2003)004<1119:annatr>2.0.co;2. [Online]. Available: https://doi.org/10.1175/1525-7541(2003)004%3C1119:annatr%3E2.0.co;2.

[18] I. Urbich, J. Bendix, and R. Müller, "A novel approach for the short-term forecast of the effective cloud albedo," *Remote Sensing*, vol. 10, no. 6, p. 955, Jun. 2018. DOI: 10.3390/rs10060955. [Online]. Available: https://doi.org/10.3390/rs10060955.

[19] X. Zhang, M. D. Goldberg, and Y. Yu, "Prototype for monitoring and forecasting fall foliage coloration in real time from satellite data," *Agricultural and Forest Meteorology*, vol. 158-159, pp. 21–29, Jun. 2012. DOI: 10.1016/j.agrformet.2012.01.013. [Online]. Available: https://doi.org/10.1016/j.agrformet.2012.01.013.

[20] J. L. Musuuza, D. Gustafsson, R. Pimentel, L. Crochemore, and I. Pechlivanidis, "Impact of satellite and in situ data assimilation on hydrological predictions," *Remote Sensing*, vol. 12, no. 5, p. 811, Mar. 2020. DOI: 10.3390/rs12050811. [Online]. Available: https://doi.org/10.3390/rs12050811.

[21] M. Hrachowitz, H. Savenije, G. Blöschl, *et al.*, "A decade of predictions in ungauged basins (PUB)—a review," *Hydrological Sciences Journal*, vol. 58, no. 6, pp. 1198–1255, Jun. 2013. DOI: 10.1080/02626667.2013.803183. [Online]. Available: https://doi.org/10.1080/02626667.2013.803183.

[22] T. L. Du, H. Lee, D. D. Bui, *et al.*, "Streamflow prediction in "geopolitically ungauged" basins using satellite observations and regionalization at subcontinental scale," *Journal of Hydrology*, vol. 588, p. 125 016, Sep. 2020. DOI: 10.1016/j.jhydrol.2020.125016. [Online]. Available: https://doi.org/10.1016/j.jhydrol.2020.125016.

[23] A. Ruhi, M. L. Messager, and J. D. Olden, "Tracking the pulse of the earth's fresh waters," *Nature Sustainability*, vol. 1, no. 4, pp. 198–203, Apr. 2018. DOI: 10.1038/s41893-018-0047-7. [Online]. Available: https://doi.org/10.1038/s41893-018-0047-7.

[24] ; Y. S. D. K. Michele Magni; Edwin H. Sutanudjaja, *Input data for the 30 arcmin hybrid global streamflow modelling framework described in 'global streamflow modelling using process-informed machine learning' (2023). contains the location of the stations and their respective observed discharge (grdc), input meteorology, input catchment parameters and the state variables and streamflow from the uncalibrated pcr-globwb.* https://zenodo.org/record/7890583, Accessed: 2023-04-17.

[25] I. Harris, P. Jones, T. Osborn, and D. Lister, "Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 dataset," *International Journal of Climatology*, vol. 34, no. 3, pp. 623–642, May 2013. DOI: 10.1002/joc.3711. [Online]. Available: https://doi.org/10.1002/joc.3711.

[26] U. Schulzweida, *Cdo user guide*, version 2.1.0, Oct. 2022. DOI: 10.5281/zenodo.7112925. [Online]. Available: https://doi.org/10.5281/zenodo.7112925.

[27] D. Karssenberg, O. Schmitz, P. Salamon, K. de Jong, and M. F. Bierkens, "A software framework for construction of process-based stochastic spatio-temporal models and data assimilation," *Environmental*

Modelling &amp Software, vol. 25, no. 4, pp. 489–502, Apr. 2010. DOI: 10.1016/j.envsoft.2009.10.004. [Online]. Available: https://doi.org/10.1016/j.envsoft.2009.10.004.

[28] E. H. Sutanudjaja, R. van Beek, N. Wanders, *et al.*, "PCR-GLOBWB 2: A 5 arcmin global hydrological and water resources model," *Geoscientific Model Development*, vol. 11, no. 6, pp. 2429–2453, Jun. 2018. DOI: 10.5194/gmd-11-2429-2018. [Online]. Available: https://doi.org/10.5194/gmd-11-2429-2018.

[29] L. P. H. van Beek, Y. Wada, and M. F. P. Bierkens, "Global monthly water stress: 1. water balance and water availability," *Water Resources Research*, vol. 47, no. 7, Jul. 2011. DOI: 10.1029/2010wr009791. [Online]. Available: https://doi.org/10.1029/2010wr009791.

[30] Y. Wada, D. Wisser, and M. F. P. Bierkens, "Global modeling of withdrawal, allocation and consumptive use of surface water and groundwater resources," *Earth System Dynamics*, vol. 5, no. 1, pp. 15–40, Jan. 2014. DOI: 10.5194/esd-5-15-2014. [Online]. Available: https://doi.org/10.5194/esd-5-15-2014.

[31] I. de Graaf, L. van Beek, Y. Wada, and M. Bierkens, "Dynamic attribution of global water demand to surface water and groundwater resources: Effects of abstractions and return flows on river discharges," *Advances in Water Resources*, vol. 64, pp. 21–33, Feb. 2014. DOI: 10.1016/j.advwatres.2013.12.002. [Online]. Available: https://doi.org/10.1016/j.advwatres.2013.12.002.

[32] B. Tapley and C. Reigber, "The grace mission: Status and future plans," *AGU Fall Meeting Abstracts*, vol. -1, p. 02, Nov. 2001.

[33] R. Schmidt, F. Flechtner, U. Meyer, *et al.*, "Hydrological signals observed by the GRACE satellites," *Surveys in Geophysics*, vol. 29, no. 4-5, pp. 319–334, Apr. 2008. DOI: 10.1007/s10712-008-9033-3. [Online]. Available: https://doi.org/10.1007/s10712-008-9033-3.

[34] A. Güntner, "Improvement of global hydrological models using GRACE data," *Surveys in Geophysics*, vol. 29, no. 4-5, pp. 375–397, Aug. 2008. DOI: 10.1007/s10712-008-9038-y. [Online]. Available: https://doi.org/10.1007/s10712-008-9038-y.

[35] M. Hori, K. Sugiura, K. Kobayashi, *et al.*, "A 38-year (1978–2015) northern hemisphere daily snow cover extent product derived using consistent objective criteria from satellite-borne optical sensors," *Remote Sensing of Environment*, vol. 191, pp. 402–418, Mar. 2017. DOI: 10.1016/j.rse.2017.01.023. [Online]. Available: https://doi.org/10.1016/j.rse.2017.01.023.

[36] W. Dorigo, W. Wagner, C. Albergel, *et al.*, "ESA CCI soil moisture for improved earth system understanding: State-of-the art and future directions," *Remote Sensing of Environment*, vol. 203, pp. 185–215, Dec. 2017. DOI: 10.1016/j.rse.2017.07.001. [Online]. Available: https://doi.org/10.1016/j.rse.2017.07.001.

[37] F. Ulaby, M. Moore, and A. Fung, "Microwave remote sensing, active and passive: Volume ii, radar remote sensing and surface scattering and emission theory," vol. 2, p. 634, 1982.

[38] R. M. Parinussa, A. G. C. A. Meesters, Y. Y. Liu, W. Dorigo, W. Wagner, and R. A. M. de Jeu, "Error estimates for near-real-time satellite soil moisture as derived from the land parameter retrieval model," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 4, pp. 779–783, Jul. 2011. DOI: 10.1109/lgrs.2011.2114872. [Online]. Available: https://doi.org/10.1109/lgrs.2011.2114872.

[39] M. Rodell and J. S. Famiglietti, "An analysis of terrestrial water storage variations in illinois with implications for the gravity recovery and climate experiment (GRACE)," *Water Resources Research*, vol. 37, no. 5, pp. 1327–1339, May 2001. DOI: 10.1029/2000wr900306. [Online]. Available: https://doi.org/10.1029/2000wr900306.

[40] B. R. Scanlon, Z. Zhang, H. Save, *et al.*, "Global models underestimate large decadal declining and rising water storage trends relative to GRACE satellite data," *Proceedings of the National Academy of Sciences*, vol. 115, no. 6, Jan. 2018. DOI: 10.1073/pnas.1704665115. [Online]. Available: https://doi.org/10.1073/pnas.1704665115.

[41] L. Breiman, *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001. DOI: 10.1023/a:1010933404324. [Online]. Available: https://doi.org/10.1023/a:1010933404324.

[42] P. Probst, M. N. Wright, and A.-L. Boulesteix, "Hyperparameters and tuning strategies for random forest," *WIREs Data Mining and Knowledge Discovery*, vol. 9, no. 3, Jan. 2019. DOI: 10.1002/widm.1301. [Online]. Available: https://doi.org/10.1002/widm.1301.

[43] F. Kratzert, G. Nearing, N. Addor, *et al.*, "Caravan - a global community dataset for large-sample hydrology," *Scientific Data*, vol. 10, no. 1, Jan. 2023. DOI: 10.1038/s41597-023-01975-w. [Online]. Available: https://doi.org/10.1038/s41597-023-01975-w.

[44] M. N. Wright and A. Ziegler, "Branger/b: A fast implementation of random forests for high dimensional data in ic/i and ir/i," *Journal of Statistical Software*, vol. 77, no. 1, 2017. DOI: 10.18637/jss.v077.i01. [Online]. Available: https://doi.org/10.18637/jss.v077.i01.

[45] K. J. Archer and R. V. Kimes, "Empirical characterization of random forest variable importance measures," *Computational Statistics &amp Data Analysis*, vol. 52, no. 4, pp. 2249–2260, Jan. 2008. DOI: 10.1016/j.csda.2007.08.015. [Online]. Available: https://doi.org/10.1016/j.csda.2007.08.015.

[46] H. V. Gupta, H. Kling, K. K. Yilmaz, and G. F. Martinez, "Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling," *Journal of Hydrology*, vol. 377, no. 1-2, pp. 80–91, Oct. 2009. DOI: 10.1016/j.jhydrol.2009.08.003. [Online]. Available: https://doi.org/10.1016/j.jhydrol.2009.08.003.

[47] W. J. M. Knoben, J. E. Freer, and R. A. Woods, "Technical note: Inherent benchmark or not? comparing nash-sutcliffe and kling-gupta efficiency scores," Jul. 2019. DOI: 10.5194/hess-2019-327. [Online]. Available: https://doi.org/10.5194/hess-2019-327.

[48] L. A. D. la Fuente, M. R. Ehsani, H. V. Gupta, and L. E. Condon, "Towards interpretable LSTM-based modelling of hydrological systems," Apr. 2023. DOI: 10.5194/egusphere-2023-666. [Online]. Available: https://doi.org/10.5194/egusphere-2023-666.

[49] E. Mussumeci and F. C. Coelho, "Large-scale multivariate forecasting models for dengue - LSTM versus random forest regression," *Spatial and Spatio-temporal Epidemiology*, vol. 35, p. 100 372, Nov. 2020. DOI: 10.1016/j.sste.2020.100372. [Online]. Available: https://doi.org/10.1016/j.sste.2020.100372.

[50] B. Martens, D. G. Miralles, H. Lievens, *et al.*, "GLEAM v3: Satellite-based land evaporation and root-zone soil moisture," *Geoscientific Model Development*, vol. 10, no. 5, pp. 1903–1925, May 2017. DOI: 10.5194/gmd-10-1903-2017. [Online]. Available: https://doi.org/10.5194/gmd-10-1903-2017.

[51] L. Brocca, P. Filippucci, S. Hahn, *et al.*, "SM2rain–ASCAT (2007–2018): Global daily satellite rainfall data from ASCAT soil moisture observations," *Earth System Science Data*, vol. 11, no. 4, pp. 1583–1601, Oct. 2019. DOI: 10.5194/essd-11-1583-2019. [Online]. Available: https://doi.org/10.5194/essd-11-1583-2019.

[52] H. Ahn, K. Sun, and K. P. Kim, "Comparison of missing data imputation methods in time series forecasting," *Computers, Materials &amp Continua*, vol. 70, no. 1, pp. 767–779, 2022. DOI: 10.32604/cmc.2022.019369. [Online]. Available: https://doi.org/10.32604/cmc.2022.019369.

[53] A. Hajjem, F. Bellavance, and D. Larocque, "Mixed-effects random forest for clustered data," *Journal of Statistical Computation and Simulation*, vol. 84, no. 6, pp. 1313–1328, Nov. 2012. DOI: 10.1080/00949655.2012.741599. [Online]. Available: https://doi.org/10.1080/00949655.2012.741599.