Subseasonal forecasting of temperature and precipitation over India using machine learning approach

A Thesis

Submitted to

Indian Institute of Science Education and Research Pune

in partial fulfilment of the requirements for the

BS-MS Dual Degree Programme

By

Jadhav Prajwal Prakashrao

IISER PUNE

Indian Institute of Science Education and Research Pune

Dr. Homi Bhabha Road,

Pashan, Pune 411008, INDIA.

April, 2023

Supervisor: Dr. Sreejith O. P.

Indian Meteorological Department, Pune

Dedicated to my beloved family and friends…

Certificate

This is to certify that this dissertation entitled '**Subseasonal forecasting of temperature and precipitation over India using Machine Learning (ML) approach**' towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune, represents study/work carried out by **Jadhav Prajwal Prakashrao** at Indian Institute of Science Education and Research under the supervision of **Dr Sreejith O. P.**, Indian Meteorological Department, Pune, during the academic year June 2022 to April 2023

Dr. Sreejith O. P.

(Supervisor)

Committee:

Dr. Neena J. Mani (Co-supervisor)

Dr. Suhas Ettammal (Expert)

Declaration

I hereby declare that the matter embodied in the report entitled ‘**Subseasonal forecasting of temperature and precipitation over India using Machine Learning (ML) approach**’ are the results of the work carried out by me at the department of Earth and Climate Science, Indian Institute of Science Education and Research, Pune, under the supervision of Dr Sreejith O. P. and the same has not been submitted elsewhere for any other degree.

Jadhav Prajwal Prakashrao

ID No. 20181112

Acknowledgements

I wish to thank and express my earnest gratitude to my guide Dr Sreejith O. P., Indian Meteorological Department Pune for his encouraging and valuable suggestion, discussion and inspiring guidance throughout the year long project.

I would like to express my heartfelt thanks to Dr Sabeerali C. T. and Dr Neena Joseph Mani for providing me guidance throughout the project and giving me valuable suggestions and inputs.

I am deeply grateful to Dr Suhas Ettammal for the encouragement in difficult times.

I would like to thank my lab members: Daniel Simon, Arijeet Dutta, Pratiksha Baruah and Bharath Krishnan to help in conducting various experiments in the projects and giving valuable inputs.

I would like to thank IISER Pune and MHRD, govt. of India for providing me with computational facilities and financial assistance to carry out my computational experiments and analysis of the same.

I would like to thank my friends, Kartik Kakade, Rashmi Chapke, Kapil Chandak, Darshan Dhanajkar, Shruti Ghodke, Aadesh Wadate, Suyog Sankhe for their support and encouragement throughout the project.

Contents

List of figures

List of tables

1. Abstract

Subseasonal forecasting (SSF) is the forecasting of the weather parameters two weeks (weather timescale) to two months (seasons timescale) in advance. SSF was considered a ‘predictability desert’ as it is too long for much memory of the atmospheric initial conditions and too short for slowly varying oceanic variability to be felt sufficiently strongly. Moreover, it is a high dimensional problem as it has to consider predictors from atmosphere-land-ocean. Thus, using various parameters as predictors that capture intra-seasonal variability from these three domains, I tried to investigate the weekly forecast of temperature and precipitation at 2- week, 3-week and 4-week forecast horizon over India by a computationally inexpensive ML model-MultiLLR, which prunes out irrelevant predictors and integrates remaining predictors linearly for each target date. After integrating the MultiLLR model with existing physics based dynamical models, the forecast is found to be more skillfull by 41-57% (for temperature) and 178-401% (for precipitation) than the operational dynamical model ERFS currently used by IMD to forecast sub-seasonal climate. It has also been found that, though dynamical models forecast are more skillfull on shorter timescale (week 2), the hybrid approach of MultiLLR comprising of both dynamical model and statistical model shows higher skill of precipitation forecast on extended range time scale (week-3, week4). However, for temperature prediction, hybrid approach doesn’t give any better prediction than statistical approach.

2. Introduction

Subseasonal to seasonal (S2S) forecasting is about forecasting of weather parameters, especially temperature and precipitation beyond two weeks, but less than a season (two months) ahead. This timescale of S2S prediction lies in between the weather forecasting timescale (upto 2-week) and the climate forecasting timescale (a season or beyond). Besides major advances in both weather and climate forecasting, S2S forecasting is still a big challenge. In order to understand that, an overview of modelling of the earth system is needed.

A predictable phenomenon occurring on a particular timescale has the potential predictability of weather or climate on that timescale (Hoskins, 2013). From fig. 1, On a continuum of time scales, there are various interactions and phenomenons occurring between and within components of the earth system (atmosphere-land-ocean-cryosphere-stratosphere) on the timescale of daily, to weekly, yearly and so on. These phenomenon and components of earth system, can be used as predictability sources on the corresponding timescale.

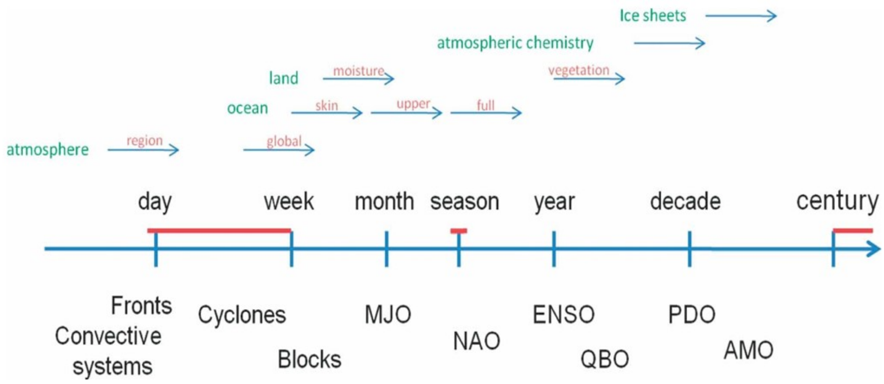

Figure 1 Time-scale is shown along the horizontal axis. Some phenomena on the different time-scales are shown at the bottom (acronyms are given in the text). At the top are indicated the components of the Earth system that need to be represented. (From Hoskins, 2013)

Based on the method of prediction, models can be classified as statistical (empirical) models and dynamical weather/climate models.

- **Statistical Models:** Statistical models use available observational data, past weather/climate patterns to find some predictive relationship between the predictors and the predictand variable. Based on the relationship found, the forecasts were made e.g., simple regression model which use past data to fit the linear model and predict the weather/climate based on best linear fit obtained.

- **Dynamical models:** In dynamical models, which are primarily numerical weather and climate prediction models, one uses governing mathematical equation representing underlying physics of the system, which on numerically solving predicts future weather or climate. Usually, the numerical evolution of equations with time is carried out over a spatial or spectral grid and the subgrid-scale processes, such as convection, are parametrized in terms of resolved scale variables.

The aim of Numerical weather prediction (NWP) is to predict the daily atmospheric state, primarily by evolving equations of the atmosphere. Other components of earth system like ocean and land surface have slow variability. Sea surface temperature (SST) and soil moisture change over a month timescale and their influence on daily weather is not deterministic (Robertson, Andrew et al., 2018). Hence, NWP becomes an initial value problem and ocean-land-cryosphere conditions are typically assumed to remain constant throughout evolution. (Lorenz, 1963) claimed that, as atmosphere is chaotic in nature, small deviations in initial conditions produces drastically different outcomes after certain timescale. He set 2-week as predictability limit for weather forecasting based on the midlatitude wave dynamics equations he considered. The tropical weather is mostly governed by convection, which has even less predictive power and in turn leads to the shorter predictability limit.

In order to get better prediction at longer timescale using NWP, better model formulation on finer grid and good forecast initialization is needed. Good forecast initialization is achieved by using data assimilation methods where observations are optimally combined with the model's previous output. By using ensemble of models with perturbed initial condition, one gets range of possible forecasts with their probabilities and uncertainty. With this advances, current NWP models has achieved significant prediction skill in extended range timescale (10-30 days), but the task requires a lot of processing power. The finer grid scale takes much memory of the system and also reduces the time step upto which skillful prediction can be achieved, thus making the NWP problem on extended range timescale computationally heavy.

Numerical climate prediction (NCP), on the other hand, is not about forecasting daily weather on a longer forecast horizon, but to predict shifts in probability distribution of climatological values of weather variables over season or beyond. Climate prediction includes processes from ocean, land surface, stratosphere, etc. along with the atmospheric component. The slow variability in SST, Soil moisture stands as good predictability source for NCP. Thus, NCP are based on same governing equations of atmosphere as the NWP but one also need to provide boundary conditions with these equations which includes effect of other components of the earth system (Robertson, Andrew et al., 2018). As the aim is to predict on larger timescale, coarser grid resolution is implemented in the model and effect of subscale processes are averaged out on climate timescale. This in turn leads to the effect of climate components to be too small to have significant impact on the prediction of daily weather patterns on a shorter timescale (Robertson, Andrew et al., 2018).

So, from the above discussion, S2S is a long timescale for much memory of the atmospheric initial condition in NWP model and too short for slowly varying boundary conditions too be felt sufficiently strongly. Prediction of weather parameters as average on long timescale and coarser grid resolution in NCP models makes the NCP forecast not useful for predicting regional S2S weather. For statistical models, S2S is a high-dimensional problem as one must consider predictors from global climate variables (SST, Geopotential height, soil moisture, etc.) and also from local weather variables (temperature, precipitation, etc.). Due of this, the S2S prediction was for long years thought of as a "predictability desert".

In recent years, there has been increase in international efforts to improve S2S forecasting. It has been spurred by the growing realization of seamless prediction across timescale. Planetary scale climate phenomenon acts as background for smaller scale features, and combined effects of smaller scale disturbances are reflected in the former one. The feedback loop between these two scales occurs on S2S timescale. 2-week predictability limit set by Lorenz (Lorenz, 1963), has been pushed forward in extended range timescale by introducing predictability source on longer timescale (like MJO). Improved understanding of S2S predictability sources like MJO, soil moisture, ocean, etc. and better forecasting methods has led to improved forecast skill in the S2S timescale.

Subseasonal timescale is an important timescale and has potential user pool in agriculture sector, water management, aviation industry, etc (White, C. J. et al., 2017). Moreover, sub-seasonal variability (corresponding to 20-90 days) accounts for the substantial portion of the Indian summer monsoon (ISM) on which India's economy is largely dependent. The modes corresponding to intra-seasonal variability are known as monsoon intra-seasonal oscillations (MISO) (Sikka and Gadgil 1980). The northward propagation of MISO plays important role in determining the onset of ISM. The active phases (above seasonal mean rainfall) and break phases of monsoon (below seasonal anomaly rainfall) are often considered as manifestation of MISO. Frequency and duration of these active-break episodes determines seasonal mean rainfall associated with ISM (Goswami and Ajayamohan 2001). Thus, prediction of MISOs, active break periods few weeks in advance has high societal importance for agriculture, disaster management, etc. across Indian subcontinent.

(Abhilash, S. et al., 2014a) attempts to predict the MISOs using ensemble of dynamical CFSv2 model, where they found general characteristics of ISM in simulations and skilful predictions till nearly 17 days. Dynamical studies have also been done for predicting active-break periods (Abhilash, S. et. al 2014b, Sushmitha Joseph et al., 2016). Study by (Goswami and Xavier at al., 2003; Suneet Dwivedi et al., 2006) uses statistical models to predict the active-break episodes and their duration by exploiting the potential predictability of the break phases. There are studies on prediction of MISO using linear or non-linear statistical models like (Xavier and Goswami 2007). Some MISO indices have also been proposed for extended-range MISO prediction and real-time monitoring (Suhas, E. et al., 2013; C T, Sabeerali et al., 2017). (Chen et al., 2018) attempts to predict the time-series of two MISO modes using physics constrained low-order non-linear stochastic model, and skilfully predicts MISO indices 20-50 days in advance.

(Judah Cohen et al., 2019) argues that in the last few decades, dynamical forecasting techniques were more skill full than the statistical approaches for prediction over the extended range mainly because of statistical techniques have not been updated since many years. It also highlights the need for new statistical technique approaches like machine learning for Subseasonal forecasting. The availability of ample meteorological climate data over India and high-performance computing makes it possible to apply machine learning (ML) based algorithms. A recent study (Jessica Hwang, 2019) tries to attempt the problem of Subseasonal forecast of temperature and precipitation over U.S region by applying simple ML forecasting system based on ensemble of two models: linear regression over multitask feature (predictor variable) selection criteria (MultiLLR model) and multitask k-nearest neighbour features (AutoKNN). The MultiLLR model chooses relevant features using backward stepwise criterion, which takes each out each feature out one-by-one and regress the model on the remaining predictors. AutoKNN model finds ‘k’ most similar dates in the past and the historical data of temperature or precipitation from these dates is used for the prediction. Both methods and their ensemble have shown higher skill in temperature and precipitation prediction on week3-4 average and week 5-6 average forecast horizon than debiased operational U.S. CFSv2 model. Other similar studies (Sijie He et al., 2020, Soukanya Mouatadid et al., 2021) tries to address the problem by using various model from ML and deep learning (DL) over the U. S. region. Moreover, It highlights use of ocean (indices such as El Nino) of land variables (soil moisture) being more helpful than the atmospheric variables.

Machine learning methods do not need complete understanding of the system. In this study, we have investigated the MultiLLR model discussed above on the predictors representing state of the atmosphere-land-ocean over India and evaluated the model’s performance over 2-week, 3-week and 4-week (in the extended range timescale) forecast horizon.

3. Data and methods

In order to predict on longer timescales, along with considering particular phenomenon occurring on those timescales, averaging weather variables on relevant timescale is important e.g., although daily weather is hard to predict on a season ahead timescale, seasonal aggregate of weather parameters is predictable a season ahead using NCP models. This ensures high-frequency disturbances averages out and the predictability of phenomenon occurring on longer timescale can be used up. In S2S problem, more precisely in extended time range (10-30 days), as aim is to predict on few weeks ahead, (Zhu et al., 2014) suggests averaging period of a week for forecasting a week ahead.

Problem statement:

To forecast weekly temperature and precipitation at the forecast horizon of 8-14 days ahead (i.e. average temperature/precipitation of week 2), 15-21 days ahead (i.e. avg. temperature/precipitation of week 3) and 22 to 28 days ahead (i.e. avg temperature/precipitation of week 4) over India (7.5° N-37.5° N) X (67.5° E -97.5° E), total 355 grid points.

Data:

Construction of the sub-seasonal dataset for training and validating the model is performed from a diverse collection of variables (listed below) which indicates the state of the atmosphere, land and ocean. Some of these variables are local weather variables (Temperature, Precipitation, surface, pressure, etc.), some are global climate variables (Sea surface temp, Geopotential height, etc.), and some are indices (Multivariate ENSO Index (MEI), RMM index). The Subseasonal data of spatiotemporal variables is prepared from the daily data of these variable, by taking the average measurements over the ensuing one-week period for each date starting from that date and interpolating on 1° X 1° grid over India (7.5° N-37.5° N) X (67.5° E -97.5° E). [Note: Words such as variable, predictors, features convey the same meaning and are used interchangeably]

A. Spatiotemporal variables:

1. Temperature: Variable (tmp2m)
   Data Source: Tmax, Tmin IMD daily data (1° x 1°) on target grid points over period (1979-2022). The same data source has been used to evaluate the model forecast.

$$tmp2m = \frac{t_{max} + t_{min}}{2}$$

2. Precipitation: Variable (precip)
   Data source: Rainfall IMD daily data (1° x 1°) on target grid points over period (1979-2022). The same data source has been used to evaluate the model forecast.

3. Relative humidity: Variable (rhum)
   It defines how much water vapour present in the air compared what it can hold.
   Data source: NCEP/NCAR reanalysis data (1948-2022)
   (https://psl.noaa.gov/data/gridded/data.ncep.reanalysis.html)

4. Sea level pressure: Variable (slp)
   Data source: NCEP/NCAR reanalysis data (1948-2022)
   (https://psl.noaa.gov/data/gridded/data.ncep.reanalysis.html)

5. Potential evaporation: Variable (pevpr)
   It is defined as the atmosphere's demand for the water vapour if sufficient water sources were present.
   Data source: NCEP/NCAR reanalysis data (1948-2022)
   (https://psl.noaa.gov/data/gridded/data.ncep.reanalysis.html)

6. Precipitable water: Variable (pr_wtr)
   It is defined the height of atmospheric columns achieved if all the water present in the columns precipitated down.
   Data source: NCEP/NCAR reanalysis data (1948-2022)
   (https://psl.noaa.gov/data/gridded/data.ncep.reanalysis.html)

7. Surface pressure: Variable (pres)
   Data source: NCEP/NCAR reanalysis data (1948-2022)
   (https://psl.noaa.gov/data/gridded/data.ncep.reanalysis.html)

8. Soil moisture: Variable (soilw)
   Data source: CPC soil moisture monthly data (1948-2022).
   (https://www.cpc.ncep.noaa.gov/products/Soilmst_Monitoring/US/Soilmst/Soilmst.shtml)

   The monthly data has been interpolated to obtain the daily values.

9. S2S multi-model ensemble: Variable (S2S ensemble)
   It is a collection of physics based dynamical subseasonal to seasonal forecast models from various modelling centres. Forecast and reforecast data of models JMA, ISAC, CNRM, CMA, KMA, NCEP been downloaded from ECMWF S2S: WWRP/WCRP Sub-seasonal to Seasonal Prediction Project.
   https://iridl.ldeo.columbia.edu/SOURCES/.ECMWF/.S2S/index.html?Set-Language=en

   A feature was constructed by taking equally weighted average of all models.

All the variables above have significant intra-seasonal variability. This has been checked by taking std. deviation of 20-90 day filtered time series of the daily data for each of the above variable.

B. Temporal variables:

1. Sea Surface Temperature: Variable (SST_2010_pc)
   Data source: NOAA dataset (1981-2022)
   (https://psl.noaa.gov/data/gridded/data.noaa.oisst.v2.highres.html)

   SST region selected: Indian ocean and west pacific ocean (30◦E – 150◦E)X(-30◦N – 30◦N)+ central and east equatorial pacific ocean (150◦E – 270◦E)X(-10◦N – 10◦N).
   (Xiouhua Fu et al., 2006) has shown extended predictability of MISOs by using a coupled atmosphere-ocean dynamical model which captures two-way interaction between MISO and sea surface. This highlight role of considering ocean parameter SST for prediction of precipitation. The selected region has the influence on generation and propogation of summer ISOs (Lau et al., 2011).

   After Principal Component Analysis (PCA) over all the selected regions using the scikit-learns' PCA package, first three PC components were extracted based on PC loading from 1981-2010.

   Variance captured by first 3 PCs (in per cent): [63.05, 14.36, 4.93]. Thus, first 3 components are capturing significant variability in the SST.

2. Geopotential height, U-V wind: Variable (wind_hgt_2010_pc)
   Data Source: NCEP/NCAR Reanalysis dataset (1948-2022)
   (https://psl.noaa.gov/data/gridded/data.ncep.reanalysis.html)

   Region: Over the whole globe. During boreal summer, there is drastic change in wind pattern over Indian subcontinent which grows strong and weak during active and break phases of MISOs respectively. Also, changes in geopotential height capture propagation of various atmospheric waves and disturbances. Thus, above parameters have been considered.

   After Principal Component Analysis (PCA) over all the selected regions using the scikit-learns' PCA package, first three PC components were extracted based on PC loading from 1948-2010

   Variance captured by first 3 PCs (in per cent):

   a) Geopotential height at 10 hPa: [91.2, 4.32, 0.81]

   b) Geopotential height at 100 hPa: [85.34, 6.07, 1.23]

c) Geopotential height at 500 hPa: [59.56, 7.29, 3.59]

d) Geopotential height at 850 hPa: [24.86, 15.16, 7.82]

e) Zonal U-wind at 250 hPa: [32.64, 3.36, 2.91]

f) Zonal U-wind at 925 hPa: [15.06, 5.17, 4.23]

g) Meridional V-wind at 250 hPa: [6.14, 4.46, 3.68]

h) Meridional V-wind at 925 hPa: [13.42, 6.03, 4.24]

3. Real time Multivariate MJO (RMM) Index: Variable (amplitude, phase)

Data-source:
http://www.bom.gov.au/climate/mjo/graphics/rmm.74toRealtime.txt

MJO is a dominant tropical intra-seasonal variation (as oscillation period spread over a rough range of 30-100 days), large scale, slowly eastward moving centre of deep convection, adjacent to which are regions of weak deep convection (Zhang, C. (2005)). Over the past few years, improvement in understanding and prediction of MJO has led to use of MJO as important source of sub-seasnonal predictability. The real-time daily multivariate MJO (RMM) index developed by (Wheeler and Hendon, 2004) has used here as a predictor for the problem.

4. Multivariate ENSO Index: Variable (mei)

Data source: NOAA MEI V2 (https://psl.noaa.gov/enso/mei/)

El-Nino Southern Oscillations (ENSO) are strong, irregular fluctuations in the tropical pacific through interactions between the atmosphere and oceans, affecting the global climate. MEI is scalar summary of atmospheric and oceanic variables associated with ENSO and thus a good indicator for the state of ENSO.

The monthly MEI index has been interpolated to daily resolution.

Methods:

Local Linear Regression with Multitask feature selection (MultiLLR) Model:

The MultiLLR model consists of local linear regression at each grid point fitted on only those features (predictors) chosen by multitask feature selection algorithm.

Local linear regression: For a grid point (g) and target date (t), we fit the linear model on training data chosen as the 56-day span (D) around day of the year of the target date (t) in any year. This 56-day span around target date (t) ensures the model is trained on a season of period around target date, and days from other seasons do not influence the prediction.

$$Y_{t,g} = \beta^T X_{t,g} = \beta_0 + \beta_1 x_{1_{t,g}} + \ldots + \beta_n x_{n_{t,g}}$$

Multitask backward stepwise feature selection: Feature selection is performed to create a model with optimum balance between bias-variance pair. Model with many features tend to have low bias-high variance and model with very few features have high bias-low variance. As we are dealing with many variables and thus have a high dimensional data, finding most relevant set of features will ensure optimum balance between bias and variance.

At each iteration, every feature is taken out step by step and the model is regressed on the remaining set of features separately each grid point. The candidate predictor that decreases the predictive performance, LOYOCV skill (defined as in the next paragraph), the least is considered irrelevant for prediction. The tolerance threshold is set to 0.01 so that the procedure will terminate when removing any predictors decrease the predictive performance by more than 0.01. This variable selection approach is known as 'frequentist', where features are chosen principally based on their statistical relevance to the target variable (predictand).

Cross-validation can be used for assessing the model's overall performance where we repeatedly divide data into test and training data and evaluate model's prediction on all the training tasks. In Leave One Year Out Cross-Validation (LOYOCV) skill method, we hold a year of data around the target date's day of the year (t) in each year iteratively (the test data) and the rest of the data is used for training the model (training data). For example, when forecasting weeks 3, we hold out a data from 22 days before t till 342 days after t; for week 4, we hold out data from 29 days before till 335 days after t, which ensure model is not fitted on dates too close to t. Predictive performance in LOYOCV is calculated as the average cosine similarity skill (defined in the next section) achieved over all these iterations.

The algorithm for feature selection can be summarized as follows,

$v \leftarrow$LOYOCV(X); converged=False
while not converged **do**
    $\forall x_j \in X; v_j \leftarrow$LOYOCV$(X/X_j)$
    if $v - max(v_j) < 0.01$ **then**
        $X \leftarrow X/x_j$
    else
        converged=True
    end if
end while
output set of final features X^*

Model Prediction Assessment:

Prediction assessment also termed as prediction skill, gives measure of how good model's prediction $\hat{y_t}$ are compared with the observation y_t. The cosine similarity skill between observed spatial anomaly vector $a_t = y_t - c_t$ and predicted spatial anomaly vector $\hat{a}_t = \hat{y}_t - c_t$ (where c_t is climatology of y_t over year 1981-2010) is given as (the vector y_t has the dimension of no. of grid points G),

$$\cos(\hat{a}_t, a_t) = \frac{\langle \hat{a}_t, a_t \rangle}{||\hat{a}_t||||a_t||} = \frac{\sum_{g=1}^{G}(\hat{a}_{t,g}.a_{t,g})}{\sqrt{\sum_{g=1}^{G}\hat{a}_{t,g}^2}\sqrt{\sum_{g=1}^{G}a_{t,g}^2}}$$

It takes values between [-1,1]. More similar are the vectors in orientation, higher will be the skill value. The cosine similarity skill of anomaly vectors is same as spatial correlation between observed anomalies and predicted anomalies. It is chosen as a prediction skill for the model as threshold can be set to 0.01, independent of prediction task.

Other prediction skills RMSE error between observed and predicted value are used just for plotting purposes. RMSE error between observed temporal anomaly vector a_g and predicted temporal anomaly vector $\hat{a}_g$ is given as,

$$RMSE(\hat{a}_g, a_g) = \sqrt{\frac{\sum_{t=1}^{N}(\hat{a}_{t,g} - a_{t,g})}{N}}$$

4. Results and discussion

In 2016, Indian Meteorological Department (IMD) has operationally implemented Extended Range Forecast System (ERFS) model which predicts daily temperature and precipitation up to extended range period (up to 30 days) starting from Thursday of every week. ERFS is an operational coupled model at IMD with a suit of models from CFSv2 coupled models. The Multi-model ensemble (MME) of 4 suit of models is run operationally for 32 days based on every Wednesday initial condition with 4 ensemble members (one control and 3 perturbed) for each suit. The forecast anomaly is generated by subtracting the climatology calculated from hindcast over 13 years period (2003-2015). The operational ERFS is used for many applications such as extended-range prediction of active-break spells of ISM, monsoon onset, progression, withdrawal, heat and cold waves, monitoring of MISOs and Madden-Julian Oscillations MJOs), cyclogenesis, and many other events. More details about the IMD's operational ERFS model can be found here:
(https://nwp.imd.gov.in/document_MME.pdf).

In order to compare between MultiLLR and operational ERFS model, forecast for MultiLLR are generated on the same dates on which ERFS model's forecast is available i.e., on Thursday of every week, over the evolution period (May 2019-April 2022). Also, the daily forecast of ERFS is converted to weekly forecast by taking average over week 2, week 3 and week 4, where each consecutive week is from Friday to next Thursday. ERFS forecasts from 1st January 2020-4th March 2020 are unavailable. Thus, those dates have not been considered in the evaluation period. Thus, the total no. of days considered in the evaluation period are 146. To mimic the real-time forecast, only the data available prior to the forecast date (starting from 1979 onwards) has been used for training the model.

Along with the ERFS model, S2S ensemble which is considered as one of the features for prediction, has also been used as a baseline to compare MultiLLR model's prediction. The forecast anomaly for S2S ensemble forecasts is calculated by subtracting the hindcast over 38 years of period (1981-2018) from S2S ensemble forecast.

Following experiments are performed to critically analyse the MultiLLR model's performance.

Experiment 1: MultiLLR on all features (Hybrid approach)

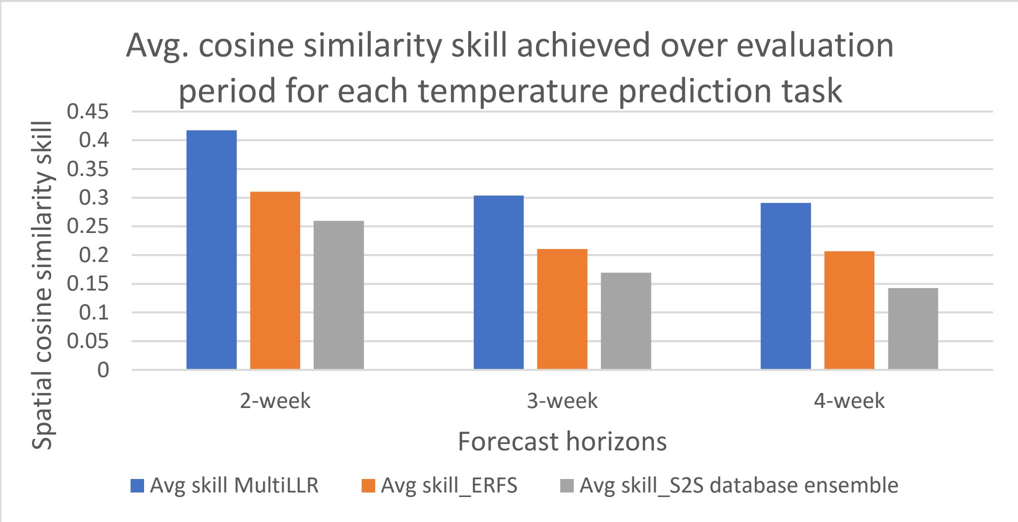

Figure 2: Average cosine similarity skill achieved over complete evaluation period for temperature forecast at various forecast horizons (2-week, 3week, 4week) by each model MultiLLR, ERFS and S2S ensemble considered

In this experiment, MultiLLR model is given all the features mentioned in the 'data' section, in the start. Figure 2 shows average cosine similarity skill achieved by MultiLLR model, ERFS model and S2S ensemble (which has also been used as one of the predictors), in each temperature prediction task (e.g., Temperature at 2week) over the whole evaluation period. We can see that MultiLLR model outperforms the baseline ERFS model, and S2S ensemble in each temperature prediction task. S2S ensemble has the least skill in each temperature prediction task than the other two models. Also, we observe that forecast skill of each model is decreasing with increase in forecast horizon.

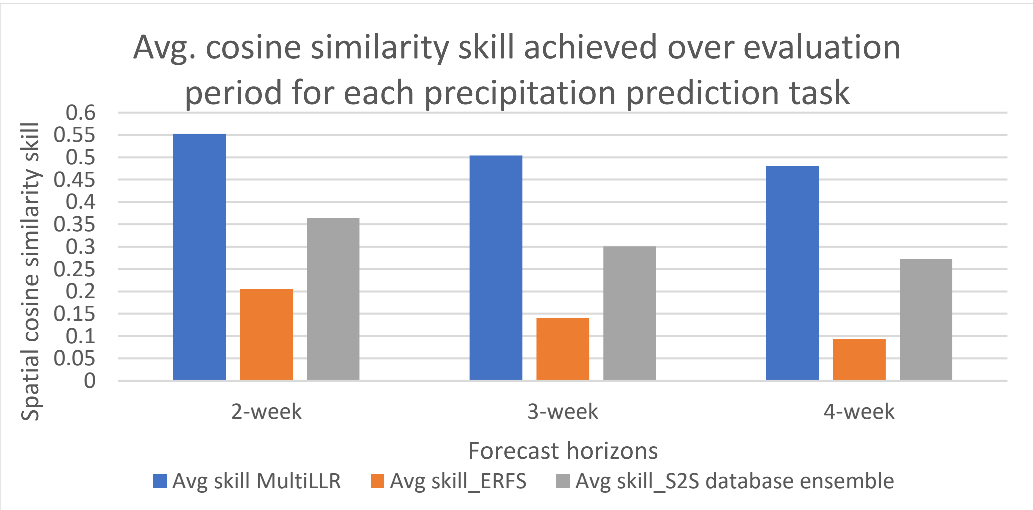

Figure 3: Average cosine similarity skill achieved over complete evaluation period for precipitation forecast at various forecast horizons (2-week, 3week, 4week) by each model MultiLLR, ERFS and S2S ensemble considered

Fig. 3, shows similar result for the precipitation prediction by each model. We can observe the average forecast skill achieved by MultiLLR model is much better than the other two models for precipitation forecast. Precipitation forecast of ERFS model has least skill compared to other two models. Also, the average skill achieved by MultiLLR over whole evaluation period decreases with increasing forecast horizon in precipitation prediction.

Table 1: Average spatial cosine similarity skill obtained by each prediction model MultiLLR, ERFS and S2S ensemble for prediction of a) temperature at 2-week b) temperature at 3-week c) temperature at 4-week d) precipitation at 2-week e) precipitation at 3-week f) precipitation at 4-week

a)

Year	Avg skill MultiLLR	Avg skill_ERFS	Avg skill_S2S database ensemble
2020	0.3957	0.2905	0.2344
2021	0.4104	0.3057	0.3055
2022	0.4463	0.3346	0.2398
Avg. over evaluation period	0.4184	0.3112	0.2617

d)

Year	Avg skill MultiLLR	Avg skill_ERFS	Avg skill_S2S database ensemble
2020	0.5522	0.2209	0.364
2021	0.5516	0.1737	0.3217
2022	0.5537	0.2206	0.4045
Avg. over evaluation period	0.5525	0.2043	0.3637

b)

Year	Avg skill MultiLLR	Avg skill_ERFS	Avg skill_S2S database ensemble
2020	0.2574	0.2188	0.1042
2021	0.3532	0.2066	0.2501
2022	0.3001	0.2062	0.1532
Avg. over evaluation period	0.3065	0.2101	0.1734

e)

Year	Avg skill MultiLLR	Avg skill_ERFS	Avg skill_S2S database ensemble
2020	0.4836	0.1195	0.2774
2021	0.513	0.1384	0.2924
2022	0.5152	0.1637	0.3324
Avg. over evaluation period	0.5051	0.1419	0.3022

c)

Year	Avg skill MultiLLR	Avg skill_ERFS	Avg skill_S2S database ensemble
2020	0.229	0.1924	0.1089
2021	0.2976	0.2324	0.2104
2022	0.3467	0.1949	0.1088
Avg. over evaluation period	0.2942	0.2076	0.1452

f)

Year	Avg skill MultiLLR	Avg skill_ERFS	Avg skill_S2S database ensemble
2020	0.4455	0.0636	0.239
2021	0.5066	0.0653	0.2753
2022	0.4883	0.1494	0.3038
Avg. over evaluation period	0.4821	0.0947	0.2748

Table 1 shows more detailed analysis of the experiment results, average skill over each year period. Here 2020 denotes a year of period from May 2019-April 2020, 2021 denotes May 2020-April 2021 and so on. Performance of MultiLLR model in each temperature and precipitation prediction task is better than other two in each year considered. The improvement in average skill over evaluation period by MultiLLR model in predicting temperature than the ERFS model is around 41% in week-2, 57% in week-3 and 57% in week-4. Similarly, the improvement in average skill by MultiLLR in predicting precipitation over ERFS is around 178% in week 2, 260% in week 3 and 401% in week 4.

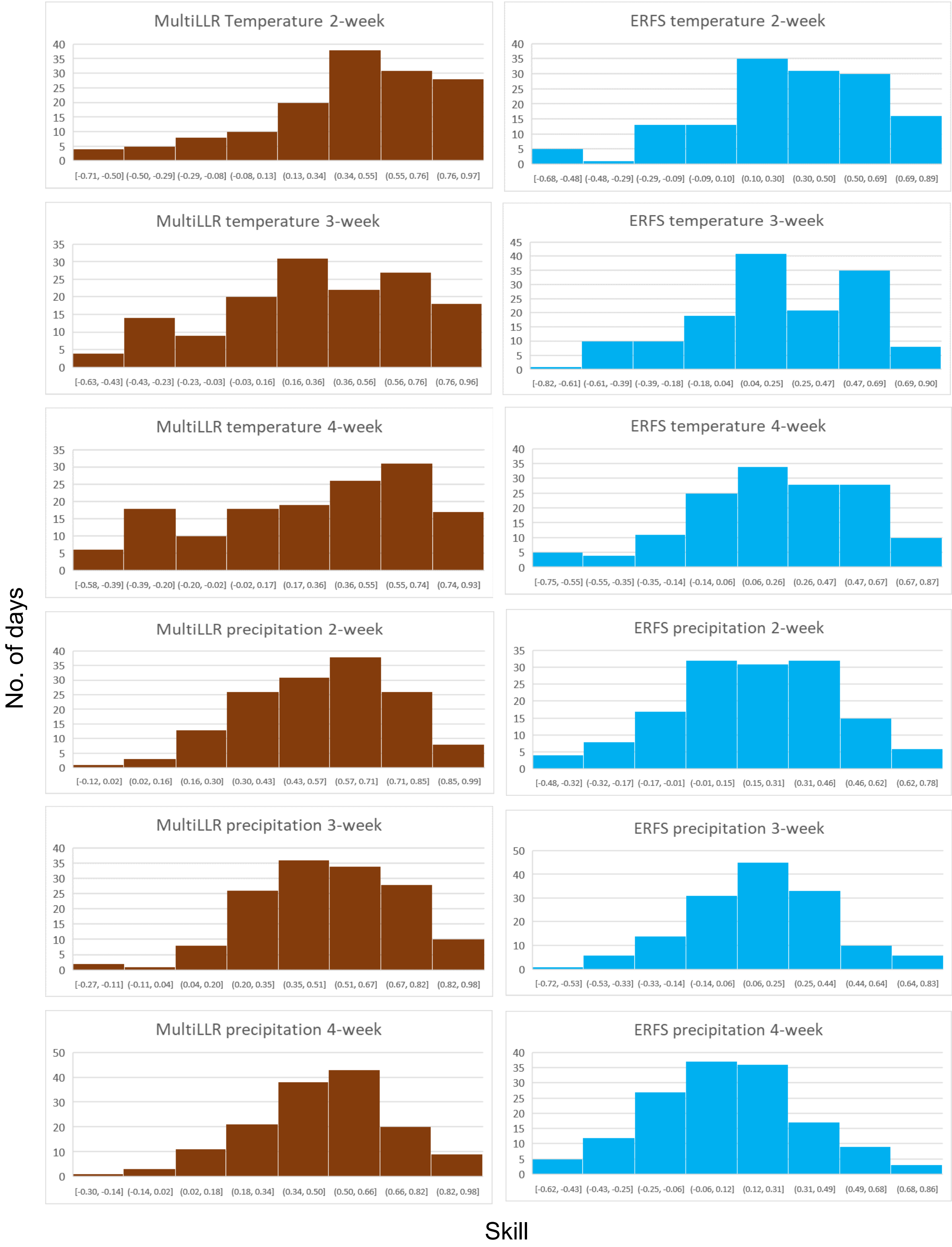

Figure 4: Skill distribution over evaluation period (146 days) for MultiLLR and ERFS model

Figure 4 shows more granular analysis of model's performance. It shows the MultiLLR skill distribution over 146 days considered in evaluation period. We see that maximum no. of days in each MultiLLR temperature and precipitation task are having positive skill with maxima in the higher skill range than the ERFS model's maxima. Though ERFS model also has positive skill on many days in each temperature prediction task, it has maximum no. of dates around '0' skill value for ERFS precipitation forecast task. This suggest MultiLLR does perform better than ERFS and can be used for operational forecasting.

Thus, in experiment 1, MultiLLR is integrating the S2S ensemble with other statistical parameters, is performing much better than S2S ensemble. This highlights the power of the hybrid approach i.e., statical models plus physics based dynamical model, for temperature and precipitation prediction is better than only using the dynamical model.

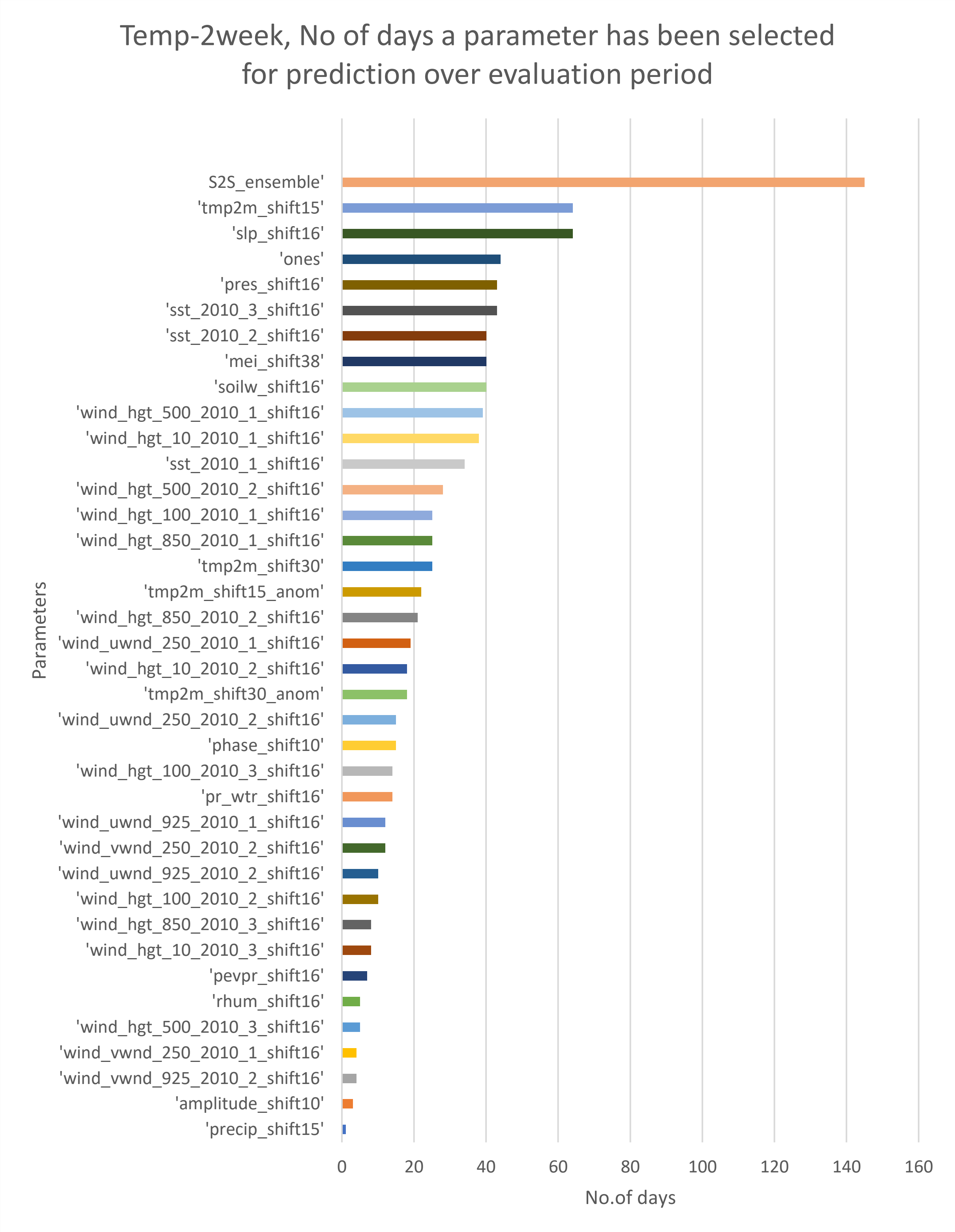

Figure 5: Feature inclusion frequency of all candidate variables for MultiLLR across all target dates over evaluation period considered for prediction of temperature at week-2 forecast horizon.

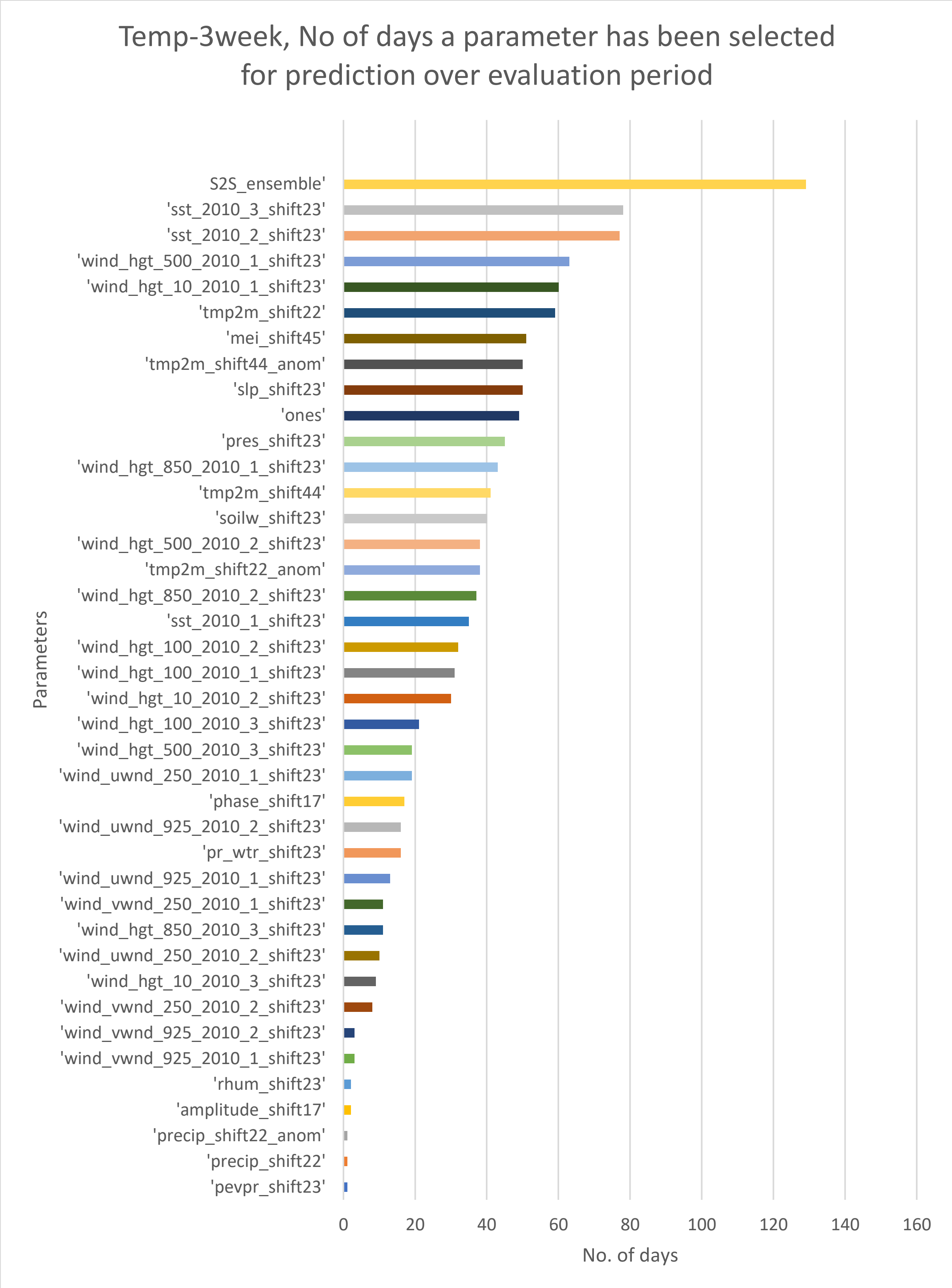

Figure 6: Feature inclusion frequency of all candidate variables for MultiLLR across all target dates over evaluation period considered for prediction of temperature at week-3 forecast horizon.

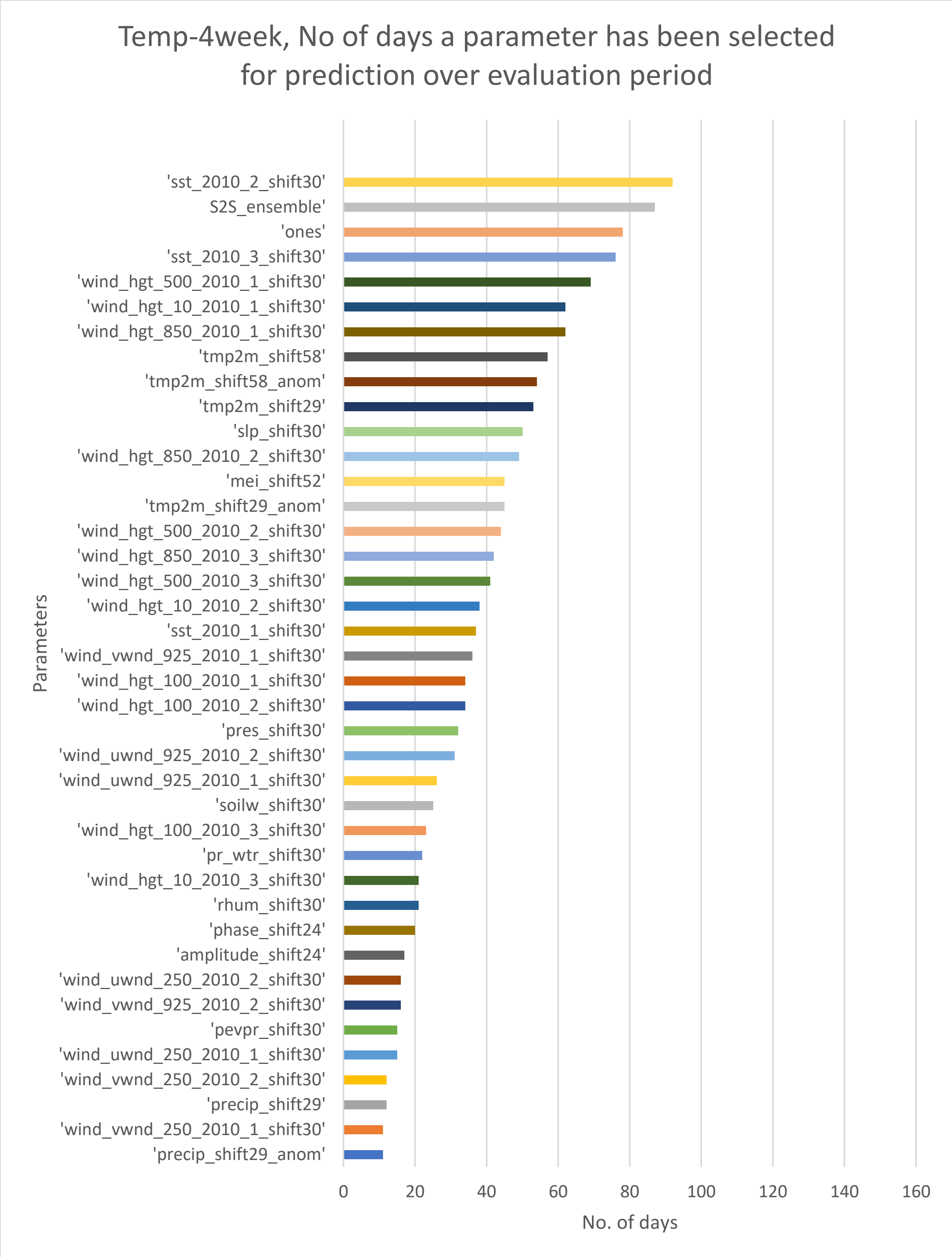

Figure 7: Feature inclusion frequency of all candidate variables for MultiLLR across all target dates over evaluation period considered for prediction of temperature at week-4 forecast horizon.

Figure 5, figure 6 and figure 7 denotes the number of days over evaluation period, a particular parameter has been selected by MultiLLR model for prediction of temperature at 2-week, 3-week, 4-week forecast horizon respectively. For temperature prediction, historical temperature values (temperature 15 days before 'tmp2m_shift15'), surface pressure ('slp_shift_16'), dynamical model forecasts (S2S ensemble) are more frequently selected by the model for prediction at 2-week and thus these parameters hold greater importance in temperature prediction. These same parameters have been relatively less frequently selected for prediction at higher forecast horizon (3- week, 4 week).

The skill of dynamical model forecasts (S2S ensemble feature here) decreases with increasing timescales. Thus, S2S ensemble feature stand as a good predictor at 2-week forecast horizon (selected for nearly every day for prediction on 2-week timescale), but is less frequently selected in 3-week and 4-week prediction task. Similarly, declining selection of historical temperature values, ('tmp2m_shift15' in 2-week, 'tmp2m_shift22' in 3-week prediction task) with the increasing forecast horizon highlights very little predictive information is shared between temperature values greater than two weeks apart.

Conversely, the relative importance of large-scale parameters like PC components of SST, geopotential height at 10 hPa and 500 hPa is increasing with the increasing forecast horizon. Geopotential height, SST are large scale climate variables and MultiLLR model has captured their importance on longer timescale.

Parameters such as precipitation, relative humidity, precipitable water, potential evaporation, MJO variables (amplitude and phase), U-V wind aren't selected by MultiLLR model frequently for temperature forecast. Multivariate ENSO Index and soil moisture are not selected much frequently but are useful for temperature prediction (selected for nearly 30-40% of dates in each prediction task). It is important to remember that the model is capturing relative importance of variable based on only their local linear dependence on temperature and thus might be failing to capture non-linear dependence of these variables on temperature. This suggests need for more complex models which will take non-linear dependence between various features and temperature into consideration.

Out of 41 no. of parameters given for training, the MultiLLR model carefully selects on an average 10,10 and 11 parameters per date for prediction of temperature.

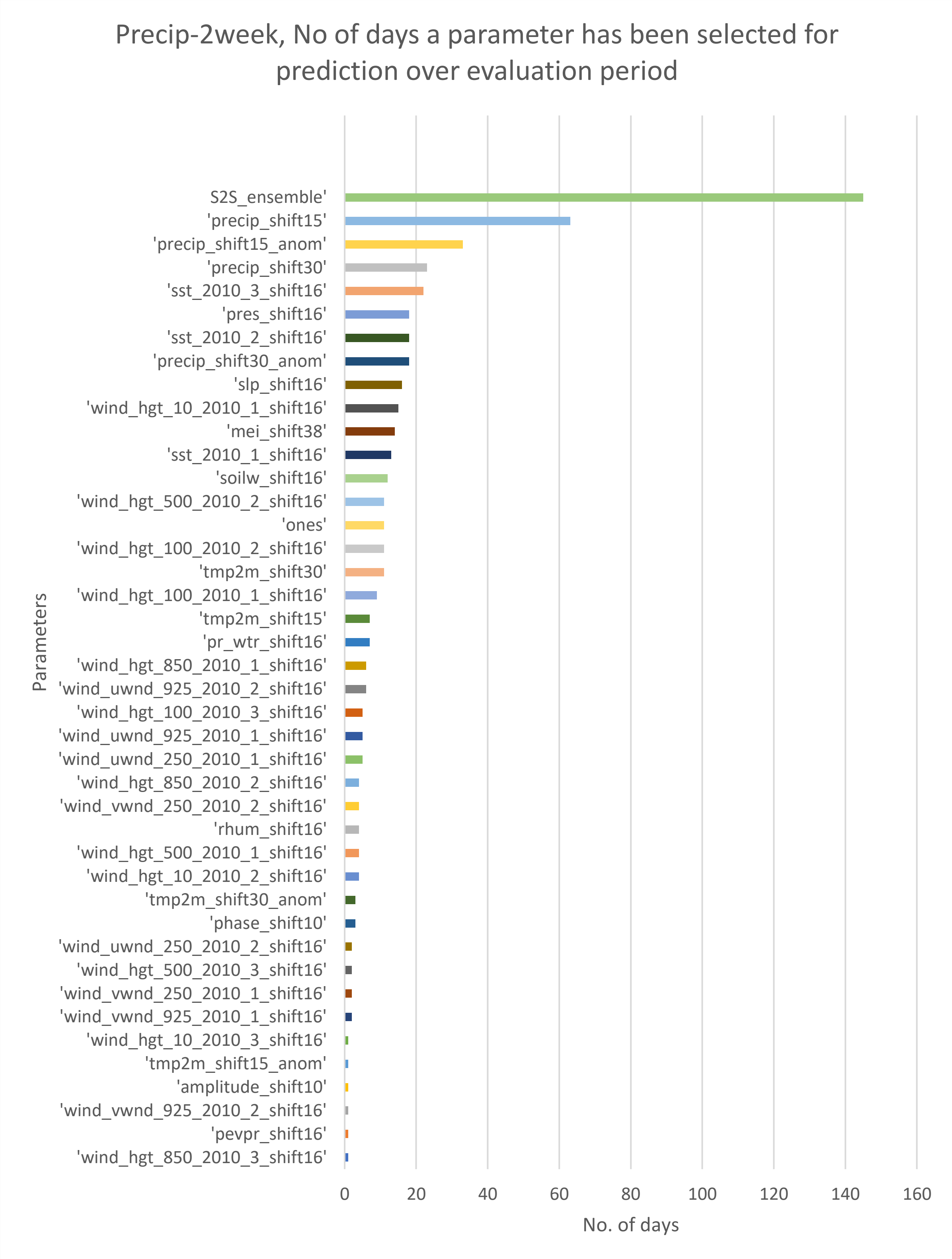

Figure 8: Feature inclusion frequency of all candidate variables for MultiLLR across all target dates over evaluation period considered for prediction of precipitation at week-2 forecast horizon.

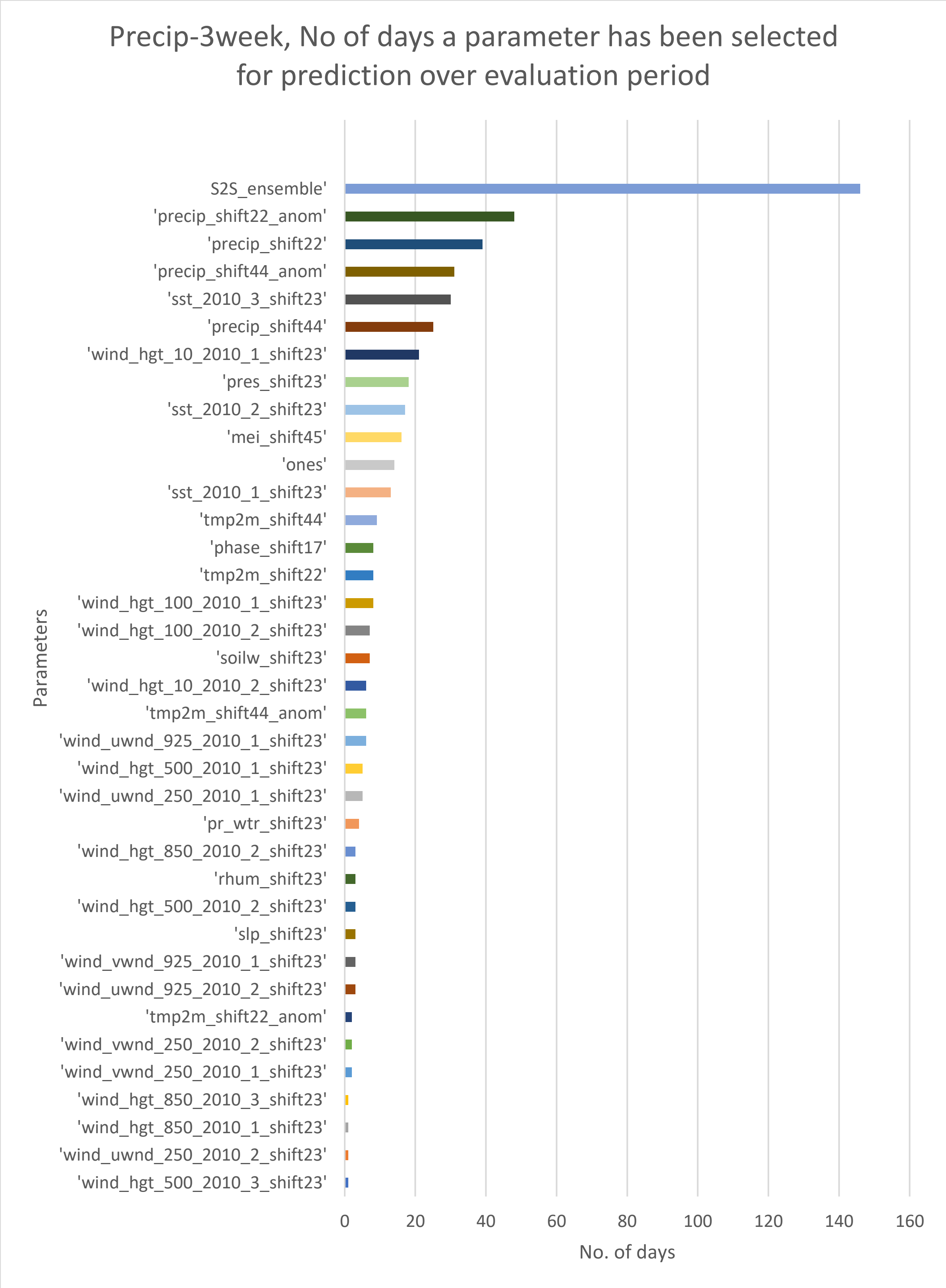

Figure 9: Feature inclusion frequency of all candidate variables for MultiLLR across all target dates over evaluation period considered for prediction of precipitation at week-3 forecast horizon.

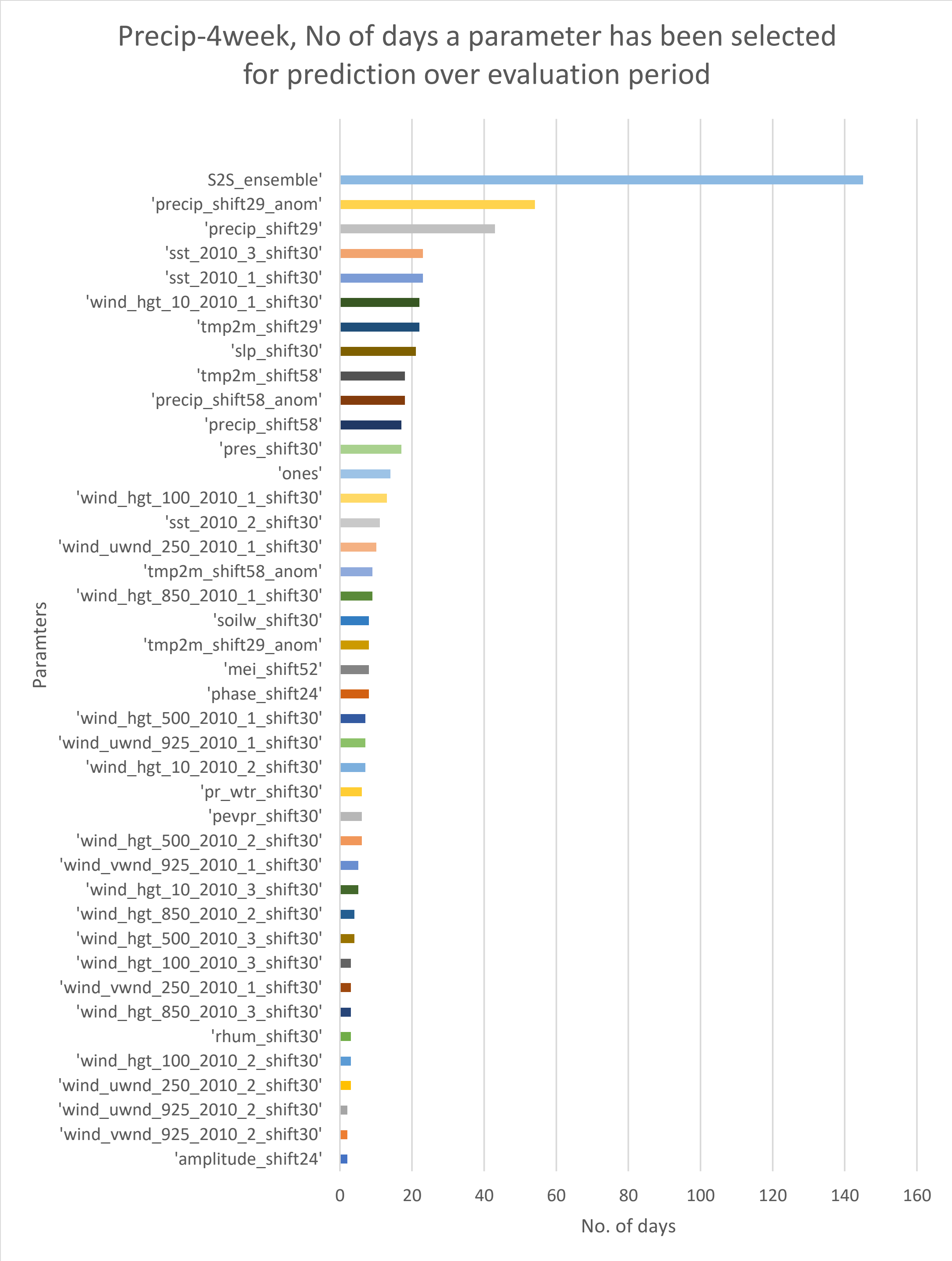

Figure 10: Feature inclusion frequency of all candidate variables for MultiLLR across all target dates over evaluation period considered for prediction of precipitation at week-4 forecast horizon.

Similar analysis can be done for precipitation task. From the figure 8, figure 9 and figure 10, it is apparent that dynamical model forecast ('S2S ensemble') is an important predictor for precipitation forecast and has been selected by nearly all the dates in the evaluation period in each precipitation prediction task. The historical observations of precipitation (e.g., ‘precip_shift15’) also seem to be more frequent parameter compared to other parameters for precipitation forecast on all the forecast horizons but its importance is decreasing with increasing forecast horizon. This highlights, with the increasing forecast horizon highlights very little predictive information is shared between precipitation values greater than 2 weeks apart.

Other parameters apart from above, do not seem to have that much relative importance and are not frequently selected by the model. Though relative importance of large-scale variables like SST, Geopotential height at 10 hPa is increasing with increasing forecast horizon, MultiLLR model has not assigned them much importance for precipitation forecast on larger forecast horizon. Relative humidity, Multivariate ENSO index, etc. are less frequently selected as predictor for prediction despite key dependence of precipitation on them. It might be due to the model is capturing relative importance of variable based only on their local linear dependence on precipitation and failing to capture non-linear dependence of these variables on precipitation.

For the precipitation task average no. of parameters selected by model for forecast at 2-week, 3-week and 4-week forecast horizon are 4,4 and 5 respectively.

Figure 11 shows the RMSE error between observation and MultiLLR temperature prediction calculated over the evaluation period. Here we see that MultiLLR model has low RMSE error over most part of the India India in general compared to the ERFS model forecast in each temperature forecast task. This highlights better prediction by MultiLLR model spatially.

One more pattern of common observation is that the RMSE error in Southern part of India is lower compared to the North India in all prediction task of MultiLLR model. It is observed that deviation in temperature over a year is higher in north India than south India and thus it was expected to get good forecast by MultiLLR model over south India. The northernmost part of India, Kashmir region, shows large error. The primary reason might be unavailability of good observational data.

Central Part of India also observe large deviation in temperature throughout the year. In ERFS model prediction tasks, there is large RMSE error over Central India. But RMSE error over same region by MultiLLR model is relatively less in each temperature prediction task.

From figure 12, the RMSE error in each precipitation forecast task by MultiLLR model has a little less RMSE error than the ERFS model in every prediction task. The region which receives highest precipitation over a year i.e., western coast of India, North East India, seems to have large RMSE in both forecasting model. Conversely, Rajasthan and rain shadow region over Deccan plateau, which receives low precipitation over a year compared to other parts of India, seems to have low values of RMSE also. Overall, MultiLLR models perform is good over India than ERFS model.

.

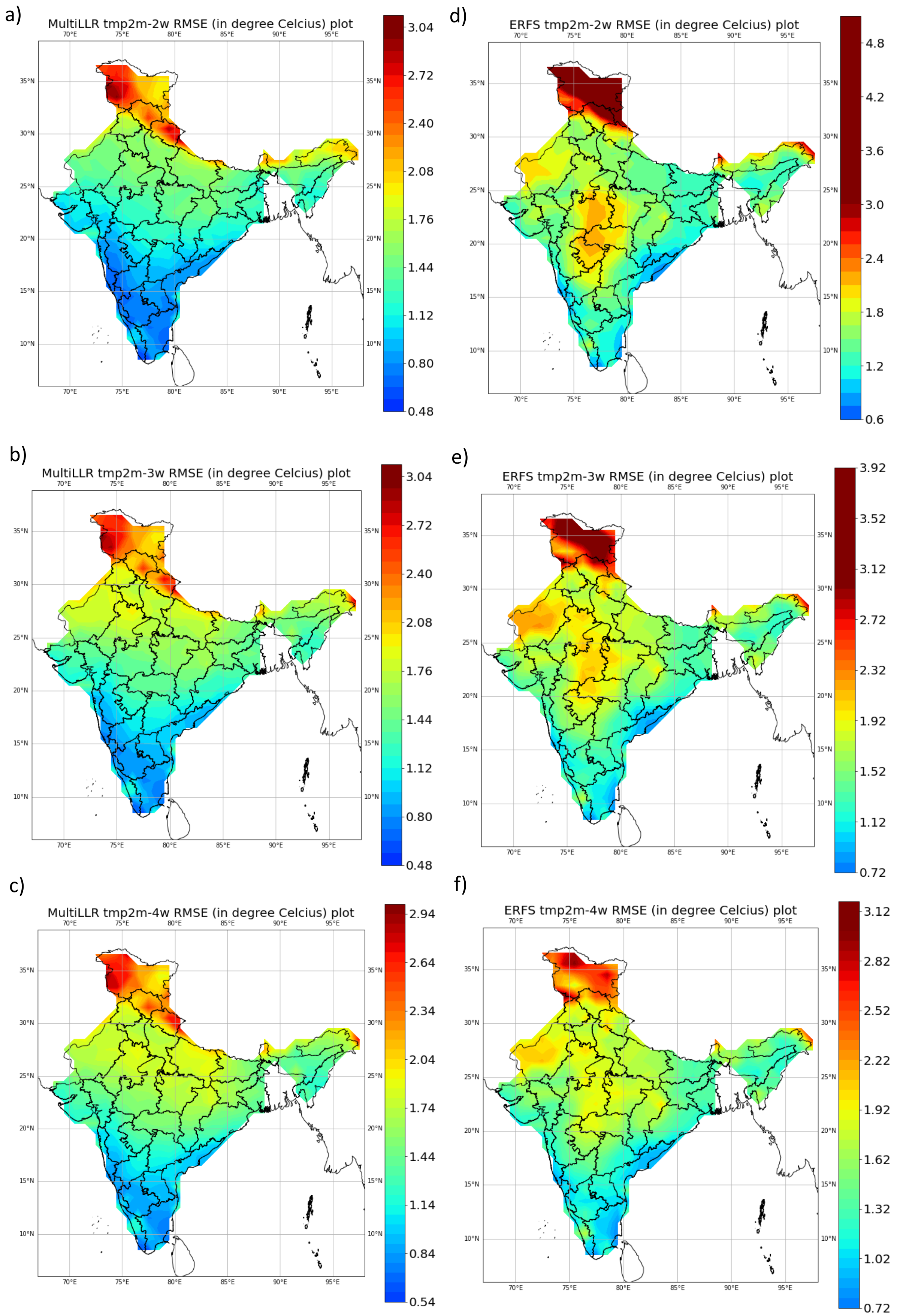

Figure 11:RMSE between temperature prediction and observation calculated over evaluation period for MultiLLR model (a), (b), (C) and for ERFS model (d), (e), (f) at 2-week, 3-week and 4-week forecast horizon respectively

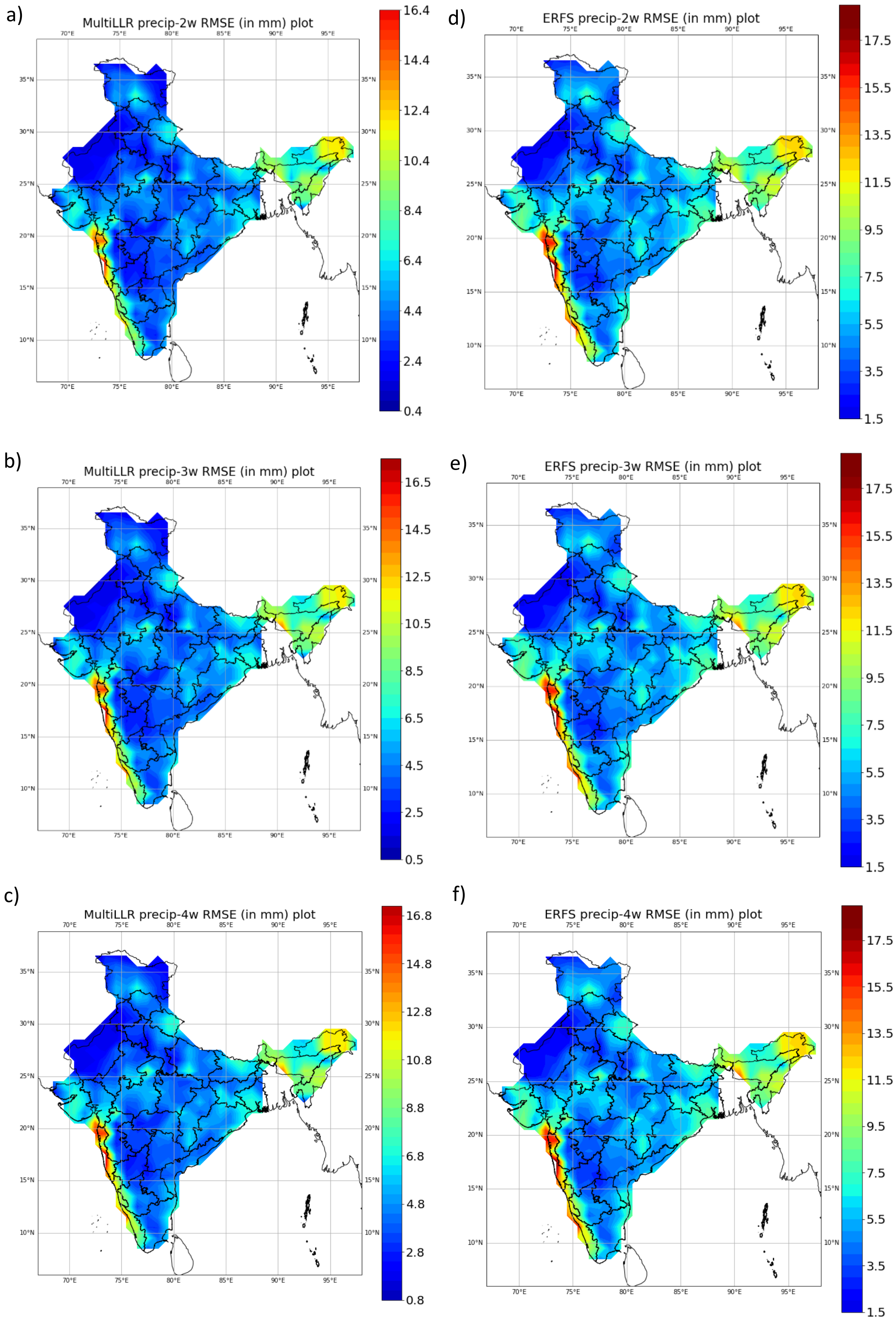

Figure 12: RMSE between precipitation prediction and observation calculated over evaluation period for MultiLLR model (a), (b), (C) and for ERFS model (d), (e), (f) at 2-week, 3-week and 4-week forecast horizon respectively

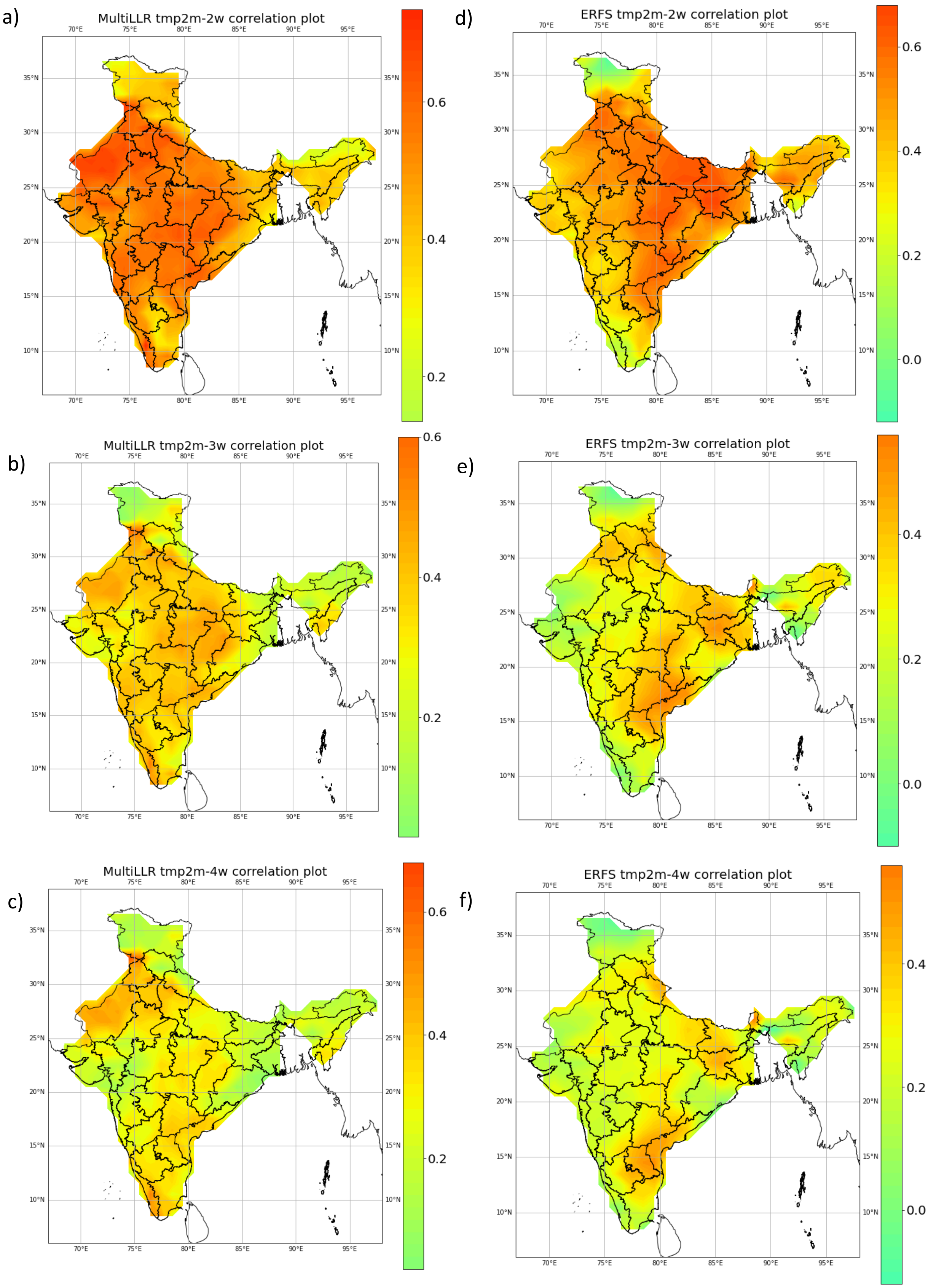

Figure 13: Correlation between temperature prediction and observation calculated over evaluation period for MultiLLR model (a), (b), (C) and for ERFS model (d), (e), (f) at 2-week, 3-week and 4-week forecast horizon respectively

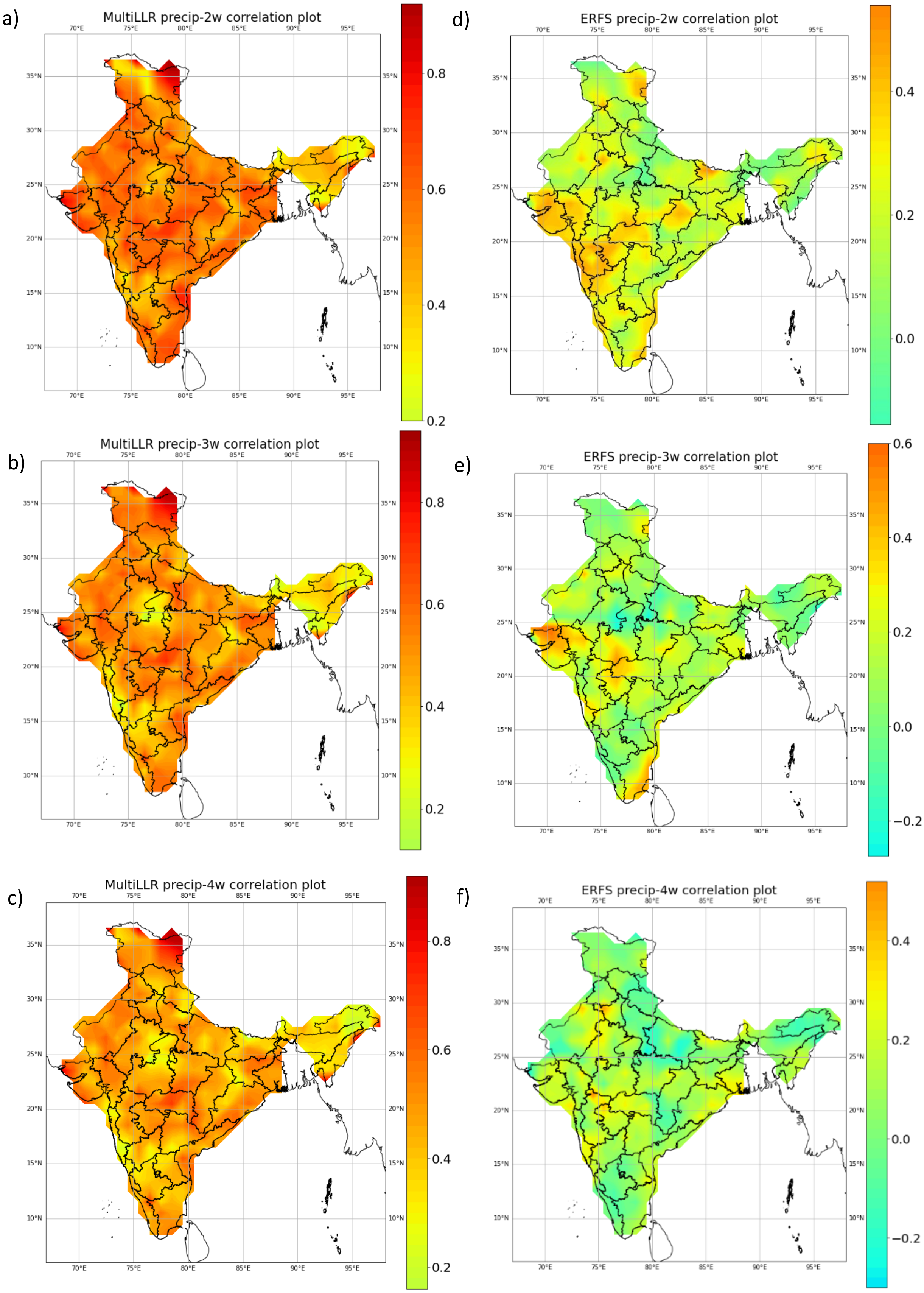

Figure 14: Correlation between precipitation prediction and observation calculated over evaluation period for MultiLLR model (a), (b), (C) and for ERFS model (d), (e), (f) at 2-week, 3-week and 4-week forecast horizon respectively

Figure 13 shows correlation between temperature forecast and observation by both models. Much part of the MultiLLR forecast region has higher correlation all over the India in each prediction task than the ERFS model.

Correlation plots for precipitation in figure 14 shows greater correlation value achieved by MultiLLR model than the ERFS model in predicting precipitation all over the forecast region.

Experiment 2: MultiLLR without S2S ensemble feature (Pure statistical approach)

S2S ensemble is one of the most selected paramters in nearly every prediction task. Thus, this experiment is carried out to see the importance of inclusion of physics based dynamical model's forecast (S2S ensemble) in MultiLLR. Experiment 2 is performed on the same evaluation period (May 2019-April 2022) and the MultiLLR model is trained on all the features except the S2S ensemble feature. The following plots show comparison between MultiLLR model results in experiment 1 (trained on all the paramters) and MultiLLR model in experiment 2.

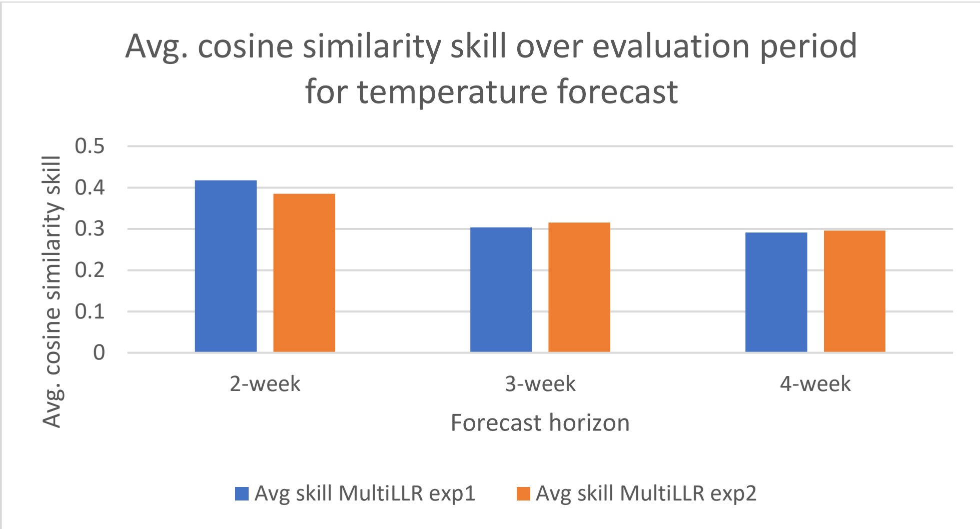

Figure 15: Comparison between average cosine similarity skill achieved for temperature prediction by MultiLLR model with(exp1) and without (exp2) S2S ensemble as a parameter

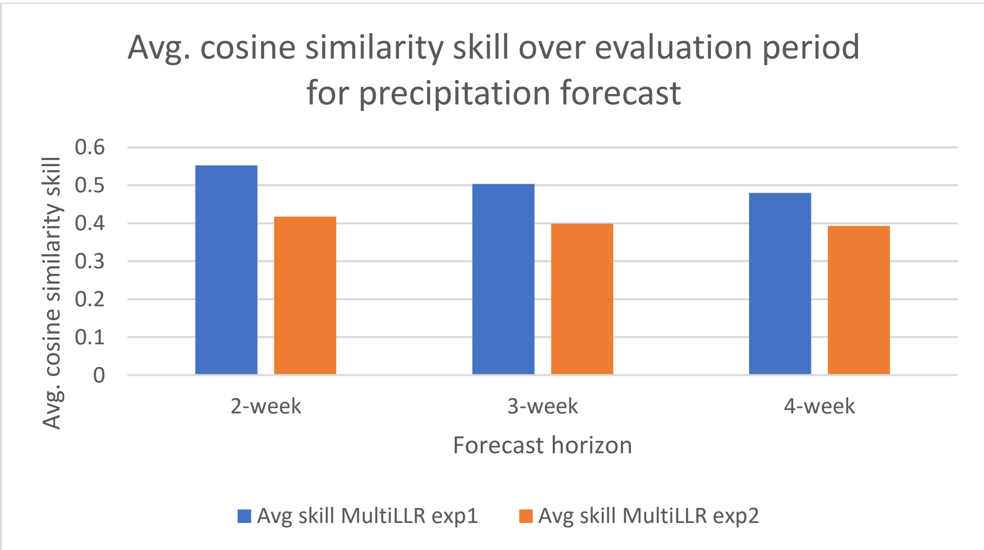

Figure 16: Comparison between average cosine similarity skill achieved for precipitation prediction by MultiLLR model with(exp1) and without (exp2) S2S ensemble as a parameter

In figure 15, we observe that MultiLLR in exp2 have greater skill than MultiLLR in exp1 only at 2-week forecast horizon, but it performs equally well as MultiLLR in exp1 on longer forecast horizon. For precipitation forecasts (figure 16), we see that MultiLLR in exp 1 has better skill compared to MultiLLR in exp 2 at all forecast horizons.

Thus, we can conclude that hybrid approach with combination of statistical parameters and physics based dynamical model's forecast is better than solely taking statistical parameters in MultiLLR for the forecast of precipitation at all forecast horizon. However, for forecast of temperature at longer forecast horizon, hybrid approach using MultiLLR model does not show any improvement.

The S2S ensemble is monthly granular and here we are trying to predict on weekly timescale. This mismatch might be leading to low performance of hybrid model in temperature forecast. It is expected that more granular S2S ensemble data will show improvement in the forecast.

Experiment 3: MultiLLR on homogenous region

The vast extent of Indian landmass leads to different regions in India experiencing varied climate. India has been divided into five homogenous regions, based on rainfall distribution and regional climatology (figure 17). The MultiLLR model in experiment 1 is selecting relevant parameters based on average forecast skill achieved over the whole forecast region (entire India) by excluding parameters one by one. As different parts of India experiences different climate at the same time, the feature which might be important for prediction in a region might not be important for forecast in another region. Thus, instead of taking whole India as a forecast region, an attempt has been made to train MultiLLR model on each of the homogenous region separately. Importance to the parameter based on avg skill achieved over the forecast region will only consider grid points falling within that homogenous region and not the entire India.

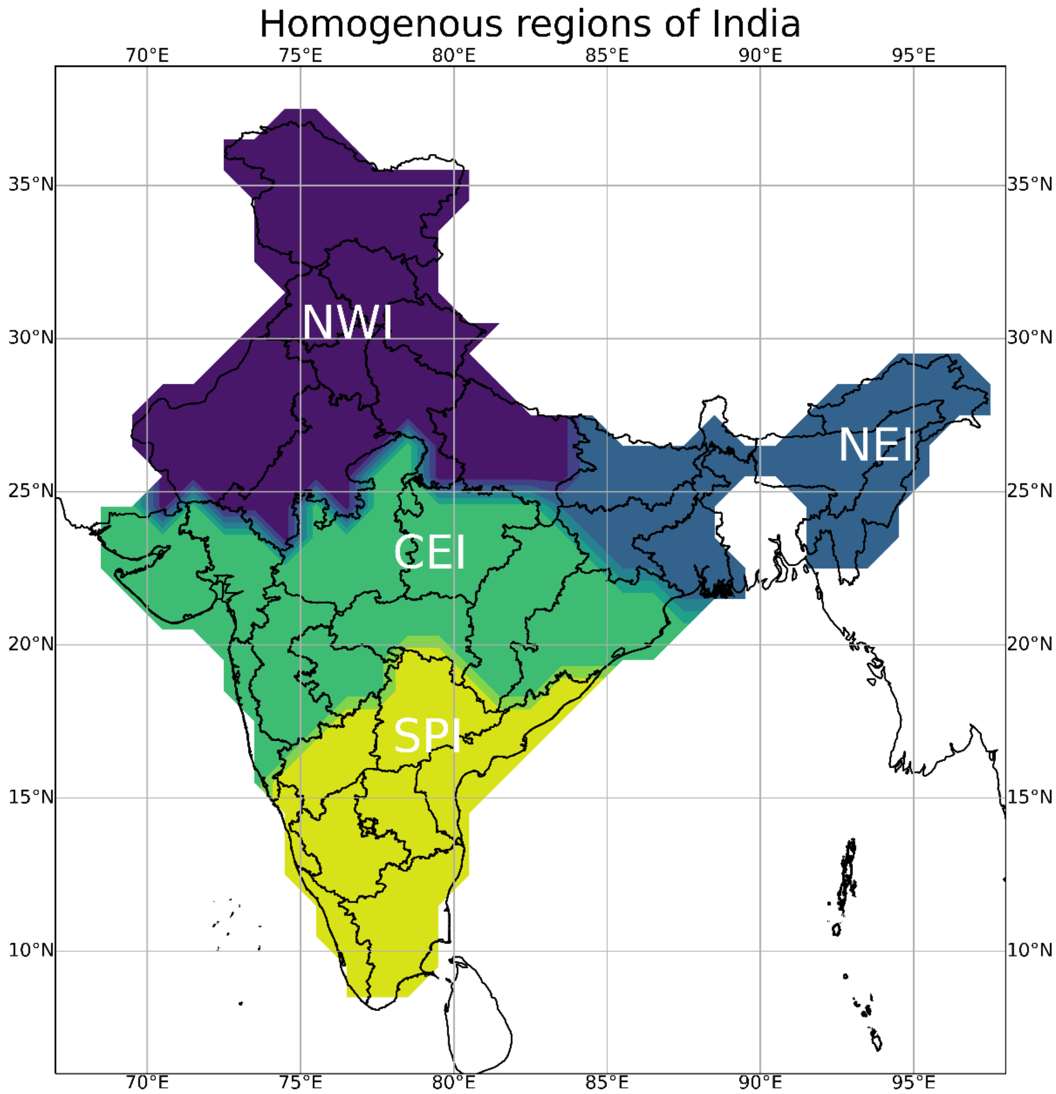

Figure 17: Four homogenous regions of India

The result of MultiLLR model in experiment 3 are summarized through following plots:

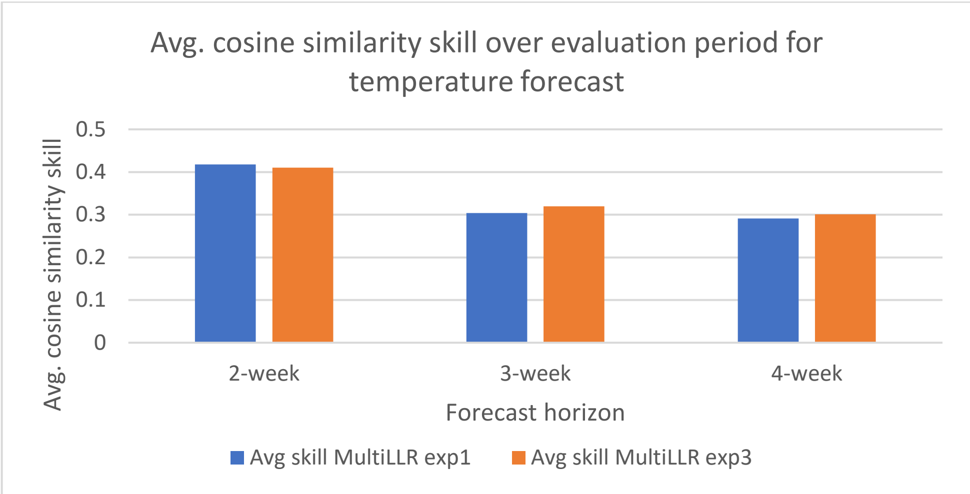

Figure 18: Comparison between average cosine similarity skill achieved for temperature prediction by MultiLLR model on entire Indian landmass(exp1) and trained on each homogeneous region separately (exp3)

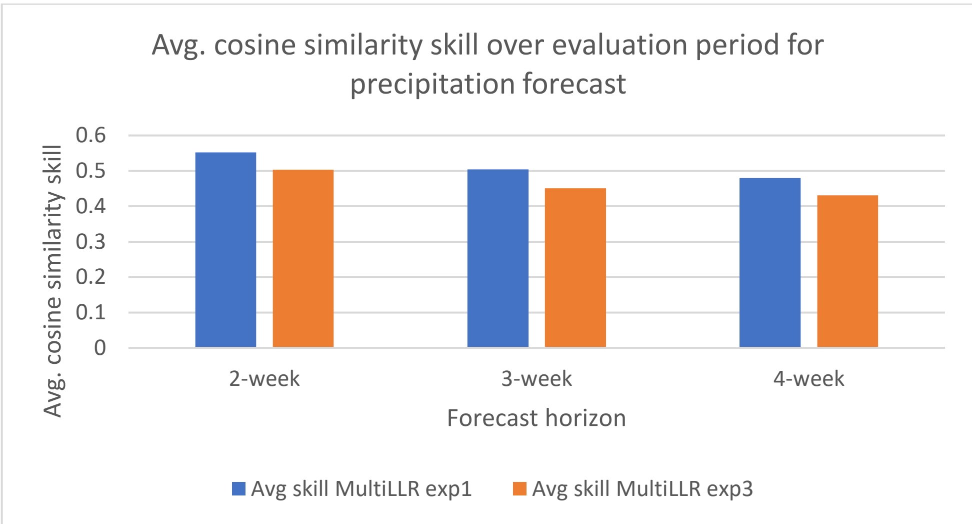

Figure 19: Comparison between average cosine similarity skill achieved for precipitation prediction by MultiLLR model on entire Indian landmass(exp1) and trained on each homogenous region separately (exp3)

It is evident from figure 18 that the approach followed in experiment 3 has shown only a slight improvement in the MultiLLR forecast average skill for week-3 and week-4 forecast horizon in the temperature forecast tasks. But besides it, there is no improvement in the average skill in any other task. For precipitation forecast in figure 19, considering entire India for feature selection seems to be more useful than training MultiLLR on each homogenous region.

It is unclear whether the homogenous region classification is not useful for the MultiLLR model or if the model doesn't perform well with less no. of grid points for feature selection.

5. Conclusion

The following conclusions can be made from the analysis carried out with MultiLLR model:

1. The average forecast skill of MultiLLR model for both temperature and precipitation is better than operational ERFS model for extended range prediction at 2-week, 3-week and 4-week forecast timescale and can be used for operational forecasting.

2. On the shorter timescale (2 week), dynamical model forecast of S2S ensemble model have good skill. Combining forecast of S2S ensemble with the other statistical parameters in the MultiLLR model shows better overall skill for precipitation forecast at all forecast horizons. Thus, for precipitation prediction, using hybrid approach is better than only statistical or dynamical model forecast. However, for temperature prediction on extended range, only statistical approach using MultiLLR model seems to have better forecast skill.

3. The model is able to capture importance of local weather parameters at short forecast horizon (2-week) and global climate parameters at longer forecast horizon (3-week and 4-week)

4. In prediction of temperature, overall performance of MultiLLR model is good over coastal part and peninsular part compared to inland part.

Simple machine learning MultiLLR model has been studied extensively and performance has been analysed on the available data. To improve the forecast skill, modification in MultiLLR like introducing non-linearity, more complex model from DL will be explored.

6. References

- Abhilash, S., Sahai, A.K., Borah, N. *et al.* Prediction and monitoring of monsoon intraseasonal oscillations over Indian monsoon region in an ensemble prediction system using CFSv2. *Clim Dyn* **42**, 2801–2815 (2014). https://doi.org/10.1007/s00382-013-2045-9

- Abhilash, S., Sahai, A.K., Pattnaik, S., Goswami, B.N. and Kumar, A. (2014), Extended range prediction of active-break spells of Indian summer monsoon rainfall using an ensemble prediction system in NCEP Climate Forecast System. Int. J. Climatol., 34: 98-113. https://doi.org/10.1002/joc.3668

- C T, Sabeerali & Ajayamohan, R. & Giannakis, Dimitrios & Majda, Andrew. (2017). Extraction and prediction of indices for monsoon intraseasonal oscillations: An approach based on nonlinear Laplacian spectral analysis. Climate Dynamics. 49. 10.1007/s00382-016-3491-y.

- Chen, N., A. J. Majda, C. T. Sabeerali, and R. S. Ajayamohan, 2018: Predicting Monsoon Intraseasonal Precipitation using a Low-Order Nonlinear Stochastic Model. *J. Climate*, **31**, 4403–4427, https://doi.org/10.1175/JCLI-D-17-0411.1

- Dwivedi, S., Mittal, A. K., and Goswami, B. N. (2006), An empirical rule for extended range prediction of duration of Indian summer monsoon breaks, *Geophys. Res. Lett.*, 33, L18801, doi:10.1029/2006GL027035

- Fu, X., B. Wang, D. E. Waliser, and L. Tao, 2007: Impact of Atmosphere–Ocean Coupling on the Predictability of Monsoon Intraseasonal Oscillations. *J. Atmos. Sci.*, **64**, 157–174, https://doi.org/10.1175/JAS3830.1.

- Goswami, B. N., and Xavier, P. K. (2003), Potential predictability and extended range prediction of Indian summer monsoon breaks, *Geophys. Res. Lett.*, 30, 1966, doi:10.1029/2003GL017810, 18

- Goswami, B.N. and Ajayamohan, R.S. (2001) Intraseasonal Oscillations and Interannual Variability of the Indian Summer Monsoon. Journal of Climate, 14, 1180-1198.

- He, S., Li, X., DelSole, T., Ravikumar, P., & Banerjee, A. (2021). Sub-Seasonal Climate Forecasting via Machine Learning: Challenges, Analysis, and Advances. *Proceedings of the AAAI Conference on Artificial Intelligence*, *35*(1), 169-177. https://doi.org/10.1609/aaai.v35i1.16090

- Hoskins, B. (2013), The potential for skill across the range of the seamless weather-climate prediction problem: a stimulus for our science. Q.J.R. Meteorol. Soc., 139: 573-584. https://doi.org/10.1002/qj.1991

- Jessica Hwang, Paulo Orenstein, Judah Cohen, Karl Pfeiffer, and Lester Mackey. Improving Subseasonal forecasting in the western us with machine learning In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2325–2335.ACM, 2019

- Joseph, S., Sahai, A.K., Chattopadhyay, R., Sharmila, S., Abhilash, S., Rajeevan, M., Mandal, R., Dey, A., Borah, N. and Phani, R. (2016), Extremes in June rainfall during the Indian summer monsoons of 2013 and 2014: observational analysis and extended-range prediction. Q.J.R. Meteorol. Soc., 142: 1276-1289. https://doi.org/10.1002/qj.2730

- Judah Cohen, Dim Coumou, Jessica Hwang, Lester Mackey, Paulo Orenstein, Sonja Totz, and Eli Tziperman. 2019. S2S reboot: An argument for greater inclusion of machine learning in subseasonal to seasonal forecasts. *Wiley Interdisciplinary Reviews: Climate Change* 10, 2 (2019), e00567

- Lau, William K-M., and Duane E. Waliser. Intraseasonal variability in the atmosphere-ocean climate system. Springer Science & Business Media, 2011.

- Lorenz, E.N., 1963. Deterministic nonperiodic flow. J. Atmos. Sci. 20 (130), 141

- Robertson, Andrew, and Frédéric Vitart, eds. Sub-seasonal to seasonal prediction: the gap between weather and climate forecasting. Elsevier, 2018.

- Sikka, D.R. and Gadgil, S. (1980) On the Maximum Cloud Zone and the ITCZ over Indian, Longitudes during the Southwest Monsoon. Monthly Weather Review, 108, 1840-1853.

- Soukanya Mouatadid et al., 2021. Learned Benchmarks for Subseasonal Forecasting. arXiv:2109.10399 [physics.ao-ph]

- Suhas, E., Neena, J.M. & Goswami, B.N. An Indian monsoon intraseasonal oscillations (MISO) index for real time monitoring and forecast verification. Clim Dyn 40, 2605–2616 (2013). https://doi.org/10.1007/s00382-012-1462-5

- Wheeler, M. C., and H. H. Hendon, 2004: An All-Season Real-Time Multivariate MJO Index: Development of an Index for Monitoring and Prediction. *Mon. Wea. Rev.*, **132**, 1917–1932,

  https://doi.org/10.1175/1520-0493(2004)132<1917:AARMMI>2.0.CO;2.

- White, C.J., Carlsen, H., Robertson, A.W., Klein, R.J.T., Lazo, J.K., Kumar, A., Vitart, F., Coughlan de Perez, E., Ray, A.J., Murray, V., Bharwani, S., MacLeod, D., James, R., Fleming, L., Morse, A.P., Eggen, B., Graham, R., Kjellström, E., Becker, E., Pegion, K.V., Holbrook, N.J., McEvoy, D., Depledge, M., Perkins-Kirkpatrick, S., Brown, T.J., Street, R., Jones, L., Remenyi, T.A., Hodgson-Johnston, I., Buontempo, C., Lamb, R., Meinke, H., Arheimer, B. and Zebiak, S.E. (2017), Potential applications of subseasonal-to-seasonal (S2S) predictions. Met. Apps, 24: 315-325. https://doi.org/10.1002/met.1654

- Xavier, P.K., Goswami, B.N., 2007. An analog method for real-time forecasting of summer monsoon subseasonal variability. Mon. Weather Rev. 135, 4149–4160. https://doi.org/10.1175/2007MWR1854.1

- Zhang, C. (2005), Madden-Julian Oscillation, Rev. Geophys., 43, RG2003, doi:10.1029/2004RG000158

- Zhu, H., M. C. Wheeler, A. H. Sobel, and D. Hudson, 2014: Seamless Precipitation Prediction Skill in the Tropics and Extratropics from a Global Model. Mon. Wea. Rev., 142, 1556–1569, https://doi.org/10.1175/MWR-D-13-00222.1.