UNIVERSITY OF CALIFORNIA,
IRVINE

Advancing Subseasonal to Seasonal Prediction Using a Composite of Models and Data

submitted as **Qualifying Proposal**
for the degree of

DOCTOR OF PHILOSOPHY

in Hydrology and Water Resources

by

Baoxiang Pan

Committee Members:
Professor Kuolin Hsu, Chair
Professor Amir AghaKouchak
Professor Soroosh Sorooshian
Professor Jinyi Yu

2017

TABLE OF CONTENTS

ABSTRACT

Advancing Subseasonal to Seasonal Prediction Using a Composite of Models and Data

By

Baoxiang Pan

University of California, Irvine, 2017

Issues about extending the accuracy and leading time of weather/climate forecasts would be explored in the thesis. The major focus is put on applying machine learning techniques, especially deep learning tools, to integrate 1)observations, 2)local high resolution simulations and 3)hindcast empirical knowledge with earth system models.

To setup baselines for the following works, the first chapter would evaluate the predictability status quo of dynamical models over a range of time scales (days to months). The focus is put on precipitation and temperature. Hindcasts from 11 operational centers' General Circulation Models that ran across 20 more years with restarting frequency of few days have been collected. Both deterministic and probabilistic scores would be applied for evaluation on daily to weekly interval temporal scales and grid to regional spatial scales.

The following chapter focuses on reducing prediction errors caused from parameterization schemes. I will explore the capacity of modern image/video processing algorithms, i.e., Convolutional Neural Network(CNN) and Capsule Networks(CapsNet), to supplement cloud microphysics parameterization schemes in estimating precipitation based on circulation constrains. Preliminary works have shown that for the West Coast United States, CNN can make good precipitation estimations using pressure data at hourly to monthly temporal scales. Jacobian of the trained CNN could be applied to explain the influence of dominating pressure's location, coverage and structure on precipitation distribution. The established connection could be used to process circulation forecast for a potentially better precipitation estimation. As an improvement to CNN that achieves state-of-the-art performance in

testing dataset, CapsNet preserves detailed information of object's location and its pose, which might contribute to even better results. Other constrains that are reported to influence precipitation, such as soil moisture and near surface air relative humidity, would be incorporated in the framework to explore land-atmosphere interactions and their roles as prediction sources for different climate regions.

The frequently restarted GCM hindcast experiments from S2S dataset offered opportunity to detect forecast windows for extended prediction. Explorative works found that during El Nino phases, precipitation predictability for California can be significantly strengthened, during La Nina phases, predictability for Oregon and Washington State can be improved. These evidences showed some possible directions for longer prediction. Firstly, some other climate background might also provide forecast windows, such as Madden Julian Oscillation and sudden stratospheric warming. Secondly, with these forecast windows detected, it is possible to design better ensemble strategies by assigning different weights to different models according to their performances under these scenarios.

The Advanced Research Weather Research and Forecasting model(WRF-ARW) is setted and used for producing local high resolution simulations. Its pressure profile output could be applied for diagnosis of the constructed circulation-precipitation connection. Also, a flood nowcast strategy by merging the relatively more reliable pressure predictions and remote sensing observations using a recurrent convolution network structure would be developped.

The following topics might be explored if time allows:

1. Illustrate the nature, difficulty of long term prediction with hierarchical models.

2. Quantify the relation between accuracy, extension of forecast and economic value by feeding hindcasts to water resources management models.

3. Land Surface Parameterization. The fourth chapter would explore the possibility of land surface simulation with the fast flourishing deep learning techniques, specifically, the convolution/capsule operator would be applied to detect spatial patterns and recur-

rent structure would be applied to cope with state variables holding memories. This might provide supplementary, even better boundary conditions to promote weather prediction over longer temporal range. Twenty years' hourly input and output data across CONUS of 3 existed land surface models(MOSAIC, VIC and NOAALSM) have been collected for model pretraining and testing. Further model tuning and verification would use FLUXNET data and some other observations.

4. Precipitation nowcast by merging remote sensing observations and numerical model outputs.

5. Explore the application of generative machine learning models in understanding ensemble weather forecasts.

Chapter 1

Introduction

1.1 Background

The accuracy and extension of weather/climate forecast is of consistent concern to humanity for its impacts on agriculture, water resources allocation, emergency management, transportation, etc.. It is necessary to keep reviewing and exploring predictability across temporal scales to exert the potential gained from model and observation advances[1, 2].

1.1.1 Lessons form Dynamical Modelling

The possibility to "calculate" weather roots in the fact that equations based on mass continuity, conservation of momentum, the first and second laws of thermodynamics, and the ideal gas law deterministically describe the state of the atmosphere, as defined by its pressure, temperature, density, humidity, and the three components of the flow velocity vector[3]. Based on this, the illustration of physical laws, improvements in observation techniques, together with the exploding computing power, have nurtured a quiet revolution in weather prediction through the past century, which stretched its coverage from several hours to beyond one week[4, 5, 6]. Below I listed the milestones of observation, theory and modelling innovations.

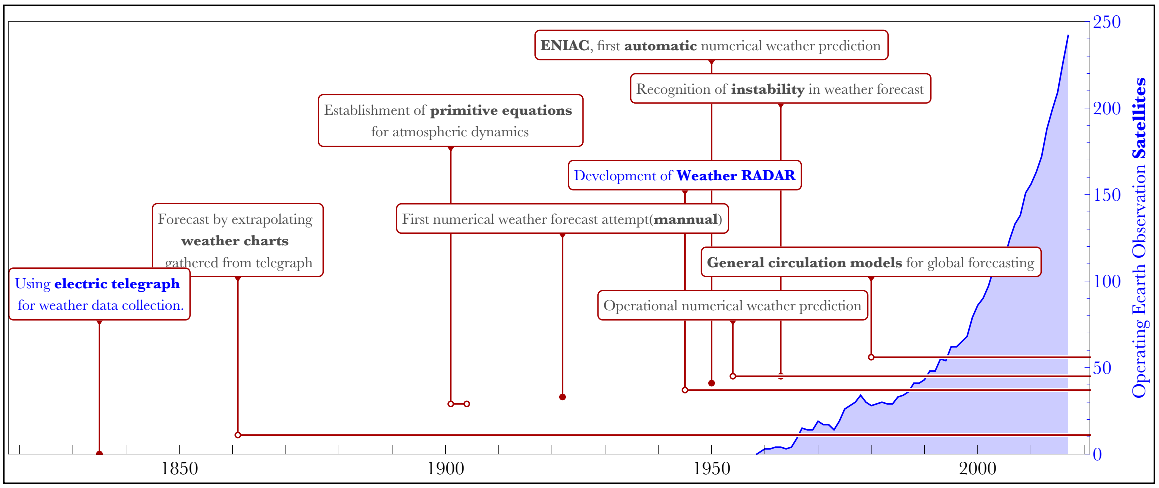


Figure 1.1: Milestones in the Development of Dynamical Weather/Climate Models. Blue parts indicates progresses in observation. Black parts indicates progresses in theory and modelling.

Corresponding to these achievements, predictability regarding 500 hpa Geo Potential Height(GPH) form European Centre for Medium-Range Weather Forecasts(ECMWF) and National Centers for Environmental Prediction (NCEP) models were listed below.

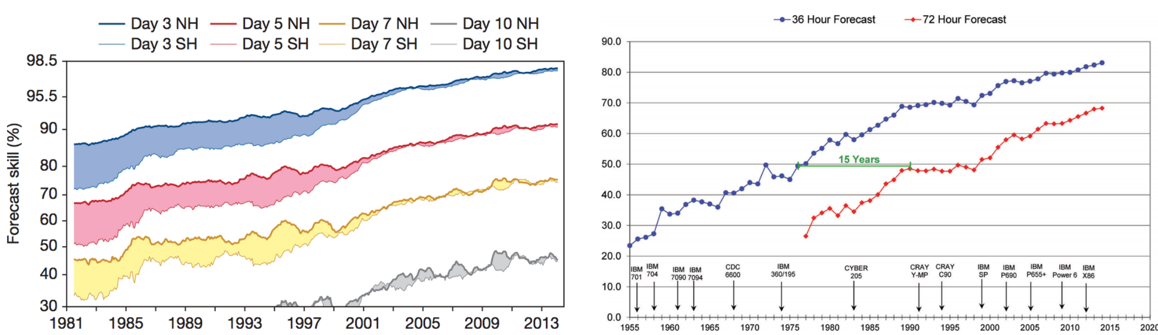


Figure 1.2: Evolution of forecast skills of ECMWF(left) and NCEP(right) model. For ECMWF[7], bold/normal line represent correlation coefficient of Northern/Southern Hemisphere 500 hpa GPH. For NCEP[8], dotted lines represent normalized mean absolute error of 500 hpa GPH prediction over North America

The advances are achieved within a classical procedure for dynamical system prediction: firstly a diagnostic step is applied to process data for initial condition estimation, which is then advanced in time by a prognostic step that solves the initial value problem[9]. Within this framework, predictability is determined by the following aspects:

1. The acceptable threshold regarding certain skill metrics at certain scales.
2. The accuracy of initial state measurement.
3. The characteristic time scale depending on the weather dynamics, known as its "Lyapunov time".
4. The accuracy and comprehensiveness of model's constitutional functions.
5. The accuracy and efficiency of model's numerical solvers.
6. If the focus is put on variables resulted from unresolved scale processes(for instance, precipitation), predictability also depends on the capacity to infer these subsidiary variables from the resolved dynamics on the computational grid[10].

Below I apply the *Lorenz System* to illustrate this point. The *Lorenz System* is an ordinary differential equation set that relates the properties of a two-dimensional fluid layer uniformly warmed from below and cooled from above[11]. Its equations and numerical simulation results were shown as follows.

$$\begin{aligned} \frac{dx}{dt} &= \sigma(y - x) \\ \frac{dy}{dt} &= x(\rho - z) - y \\ \frac{dz}{dt} &= xy - \beta z \end{aligned} \tag{1.1}$$

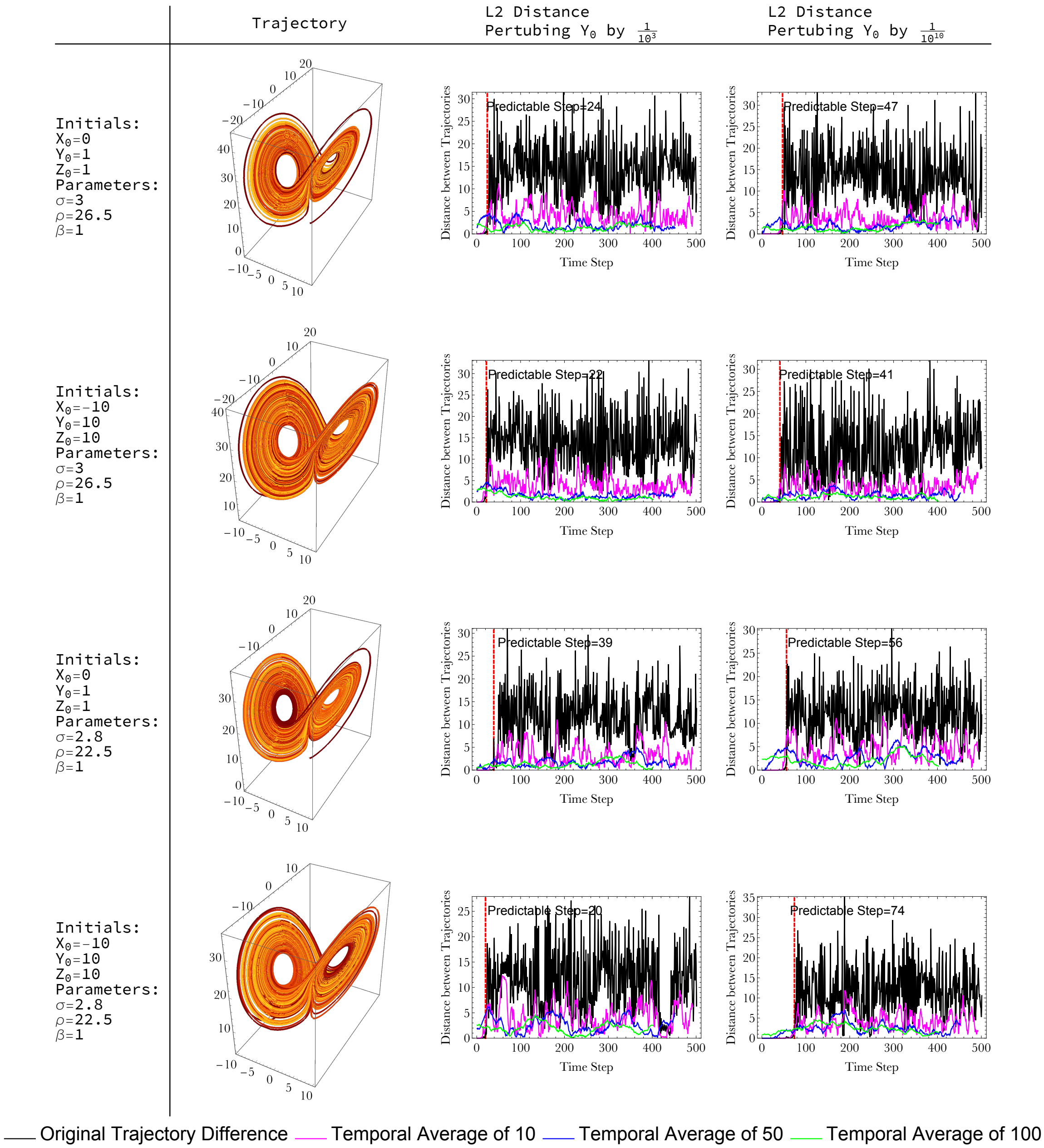


Figure 1.3: Numerical simulation results of *Lorenz System* from different initial and parameter settings. The first column represent 3-D trajectory. The last two columns shows the Euclidean Distances between trajectories due pertubation of Y_0. Dashed lines indicate predictable step, defined as smaller than 5% of the largest deviation. Colored lines depict distances between two trajectories after temporal mean filtering.

In the Euclidean Distance sense, all the cases showed trajectory divergence after certain computing steps. When the evaluation is performed on temporal average scales, divergences become less significant, given the existence of two attractors shown in Column 1. This gives some indication on the prediction of climate, i.e., how the atmosphere "behaves" over relatively long periods of time. Another way to evaluate long period performance is to compare results in the sense of their temporal distributions, which is achieved through squashing points at different time steps into a same sample space. Results of temporal distribution projected on their spatial dimensions were shown in Figure 1.4. Three sets of experiments were included. The first(black) is the temporal distribution of original trajectory, estimated using Gaussian kernel method, the second(red) uses results from a perturbed initial, the third(blue) results from adding perturbation to the Lorenz functions, this is because in weather/climate prediction, there is no universal constitutional function set that provides definite and comprehensive description of the climate system, and to illustrate this point, I add artificial perturbation of random number sampled from uniform distribution of [-0.5,0.5] to the right of each function in Equation Set 1.1.

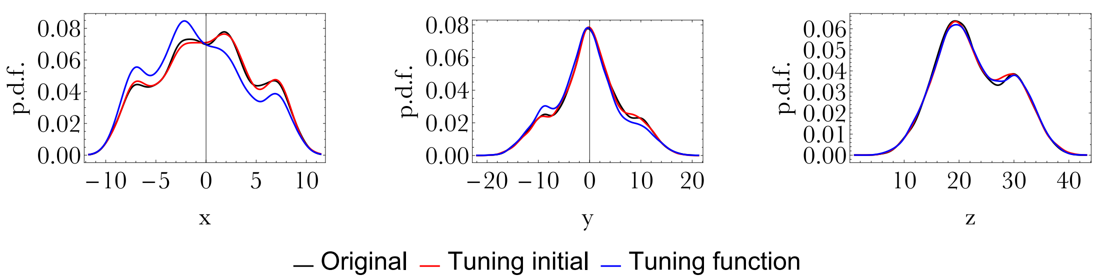


Figure 1.4: Probability distribution of trajectories projected on 3 dimensions.

Temporal distributions of trajectories starting from exact and perpetuated initials are indistinguishable regarding the three dimensions. However, function tuning influences the estimation of distribution on the X axes. Such discrepancy is named as "Hawkmoth Effect"[12], corresponding to the "Butterfly Effect"[13] that causes trajectory divergence but introduces no temporal distribution difference.

The predictable step is shown to be influenced by initial condition, magnitude of initial perturbation and function parameters. To put this argument in the weather/climate prediction context, we might expect that the "Lyapunov Time" for different climate divisions at different scenarios might be different, which offers prediction opportunity during certain prediction windows. Huge economical potential would be realized if we could recognize, explain and apply such "windows of opportunity". If hindcast empirical knowledge could be used for detection and explanation of such prediction windows,[14, 15].

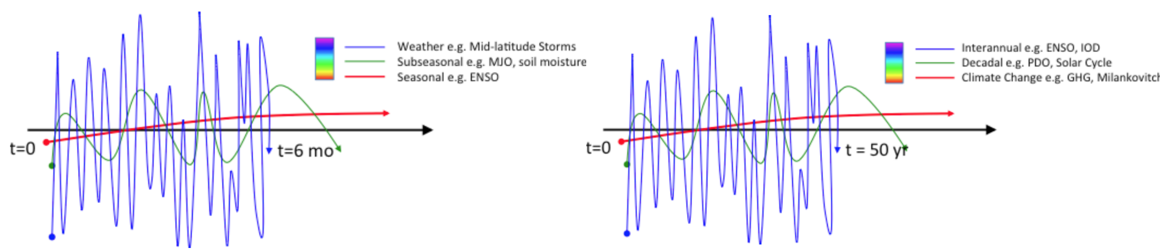


Figure 1.5: Typical Scales of different atmosphere events[15]

Although the Lorenz System here does not include implicit variables, some of them are of significance to humanity, i.e., precipitation. Generally precipitation predictive skill falls behind fundamental variables, such as temperature and pressure, which could be directly resolved in preliminary equations. As shown in Figure 1.6, a big source of prediction error for implicit variables comes from parameterization.

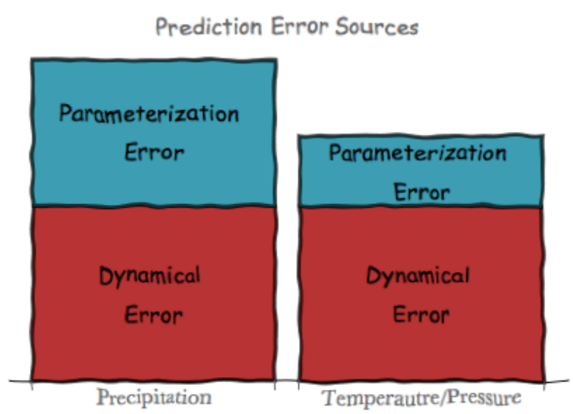


Figure 1.6: Prediction error sources for precipitation and fundamental variables.

The ratio between parameterization error and dynamical error might not be constant, for example, considering extended range prediction, as shown in Figure 1.8, If the ratio of parameterization errors remained fixed across leading time, the decreasing rate of predictive skills for precipitation and temperature/pressure would be the same. Since we usually have faster diminish of precipitation prediction skills, we infer that precipitation parameterization error increases with dynamical errors, which sets it in disadvantage on extended forecast.

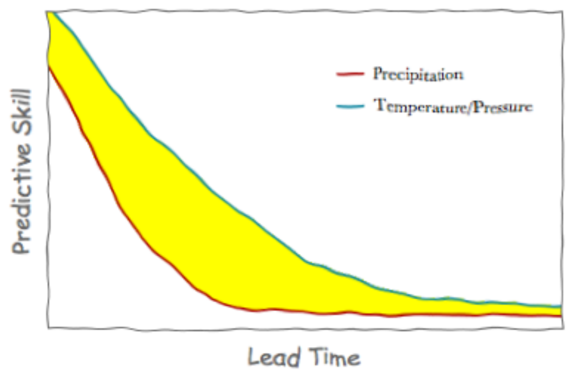


Figure 1.7: Prediction error as function of forecast leading time.

Materials above briefly reviewed the constrains against which we are striving for extended forecast, using an analog of Lorenz System. It should be pointed out that the Lorenz System might not be the perfect example here. Michael Ghil showed that a one-dimensional (1-D) atmospheric model could exhibit two steady-state evolving patterns. This might provide better illustrations[16]. I will leave it for the upcoming works.

1.1.2 Learning from Data

Advances are more likely to be generated through going back and forth between data and modelling[17]. With the accumulation of observations, we are faced with the challege of how to fully digestion them to realize their use. On the other hand, data are more and more used to guide knowledge finding processes and help making predictions, known as "Machine Learning".

The past decades have seen a booming of machine learning techniques accompanied by "infinite" available digital data and exponentially increasing computing capacity. Classical machine learning algorithms have long been introduced to the geo community to illustrate the data revealed knowledge, for instance, runoff generation simulation[18], remote sensing precipitation estimation[19], climate prediction[20], etc..

Among various learning algorithm paradigms, Deep Learning attracts most attention for its flexibility in simulating and generating big data. The figure from Andrew NG gives an intuitive explanation on the popularity of Deep Learning.

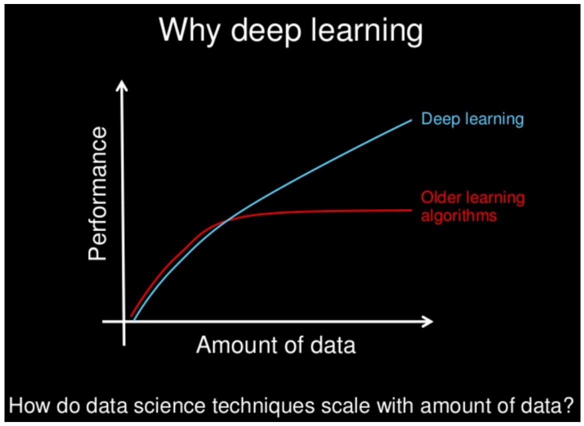


Figure 1.8: Deep Learning V.S. other machine learning paradigms[21].

The advantage of deep learning is that it is flexible enough to extract and take use of the inner structure of data for specific targets. For instance, Convolutional Neural Network(CNN) shows advantage in processing spatial structured data, which has been used for short range weather prediction[22], precipitation nowcasting[23] and extreme weather detection[24, 25], etc..For data with temporal structures, Recurrent Neural Network(RNN)[26] is preferred.

1.2 Research Questions

The fast development of observation and modelling techniques keeps offering opportunities for more exact and longer predictions through tackling the constrains clarified above. To make a comprehensive assessment would be the first step toward further progresses.

With the baselines established, issues about extending the accuracy and leading time of weather/climate forecasts would be explored, with a major focus putting on applying machine learning techniques, especially deep learning tools, to integrate 1)observations, 2)local high resolution simulations and 3)hindcast empirical knowledge with earth system models.

Focusing on extending the forecast range, the following research questions would be illustrated:

1. What is the status quo of current general circulation models in predicting key climate variables across temporal spatial scales?

2. Could machine learning help incorporating observation revealed knowledge to support better parameterization schemes in earth system models? For instance, in the case for precipitation, it is estimated with parameterization schemes resolved across computing steps(minutes scale), which is pretty sensitive to dynamical deviations, as illustrated in the previous part. Would it be possible to arrive a more coarsely resolved, but longer forecast, using its connection with resolved variables?

3. Explore the possibility of merging observations with regional circulation models for better weather forecast.

4. How to extract and apply hindcast empirical knowledge to support extended range forecast?

Chapter 2

Data and Materials

2.1 Observation Data

Observed precipitation, temperature data would be used for model construction and verification. Precipitation data were obtained from Climate Prediction Center(CPC) Unified Gauge-based precipitation products[27]. The database is constructed by merging gauge observations, satellite estimates, and numerical model predictions. It provides solid daily precipitation records covering Contiguous United States from 1948 to 2017 with spatial resolution of $0.25° \times 0.25°$. Temperature data were obtained from CPC Global Daily Temperature database[28]. It provides $0.5° \times 0.5°$ global daily temperature form 1979 to present.

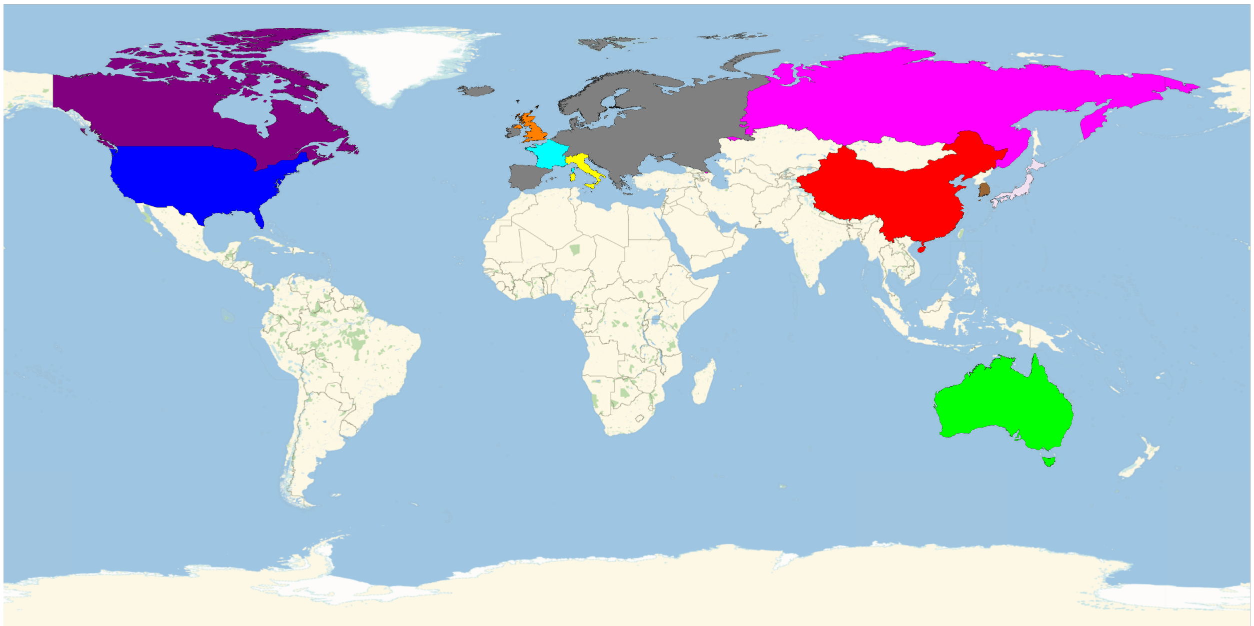
2.2 Reanalysis Data

Reanalysis data provided by the National Centers for Environmental Prediction(NCEP) and National Center for Atmospheric Research(NCAR)[29] and european Center for Medium-range Weather Forecasting(ECMWF)[30] were used. Spatial resolutions of the databases are both $2.5° \times 2.5°$ covering whole globe. Both databases provide 4 times daily records.

2.3 Hindcast Data

Hindcasts from 11 operational centers would be applied. The database is archived under the Subseasonal to Seasonal(S2S) Prediction Project science plan[31], contributed by 11 operation centers, namely the Australian Bureau of Meteorology(BoM)[32], the China Meteorological Administration(CMA)[33], the European Centre for Medium-Range Weather Forecasts (ECMWF)[34], the Environment and Climate Change Canada (ECCC)[35], the Institute of Atmospheric Sciences and Climate of the National Research Council(ISAC-CNR)[36], the Hydrometeorological Centre of Russia(HMCR)[37], the Japan Meteorological Agency(JMA)[38], the Korea Meteorological Administration (KMA)[39], Météo-France/Centre National de Recherche Meteorologiques (MétéoFrance)[40], the National Centers for Environmental Prediction (NCEP)[41], and the U.K. Met Office (UKMO)[42]. Basic information about model resolution, ensemble size, forecast extension, frequency and coupling were shown in Figure 2.1.

For each hindcast case, model is initialized with realistic estimates of their observed states, hereafter iteratively predicts the weather for a pre-setted extension without any boundary constrains. Detailed model configurations are important but beyond the coverage of the paper here. To obtain the data and acquire specific model settings, please refer to the online S2S descriptions: `https://software.ecmwf.int/wiki/display/S2S/Models`.



Model	Ensemble Size	Extension(day)	Hincast Frequency	Ocean Coupling	Sea ice Coupling
BoM(Australia)	33	62	Every 5.1 days from 1981 to 2013	Y	N
CMA(China)	4	60	Every day from 1994 to 2014	Y	Y
ECCC(Canada)	4	32	Every 4.1 days from 1995 to 2014	N	N
ECMWF(Europe)	11	46	Every 2.6 days from 1995 to 2016	Y	N
HMCR(Russia)	10	61	Every 2.9 days from 1985 to 2010	N	N
ISAC-CNR(Italy)	1	31	Every 5. days from 1981 to 2010	N	N
JMA(Japan)	5	34	Every 10.1 days from 1981 to 2010	N	N
KMA(SouthKorea)	3	60	Every 10.9 days from 1991 to 2010	Y	Y
Météo-France(France)	15	61	Every 15.2 days from 1993 to 2014	Y	Y
NCEP(UnitedStates)	4	44	Every day from 1999 to 2010	Y	Y
UKMO(UnitedKingdom)	3	60	Every 7.6 days from 1993 to 2015	Y	Y

Figure 2.1: S2S models

2.4 Climate Indices

Climate indices provide distilled information about climate scenarios, which might influence global/local circulation conditions and model's predictive skills. Here I collected most of the popular indices that quantify El Niño Southern Oscillation(ENSO) and large scale circulation patterns, such as Arctic Oscillation, Pacific Decade Oscillation and Pacific North America, etc.. The collected data were normalized and ploted in Figure 2.2

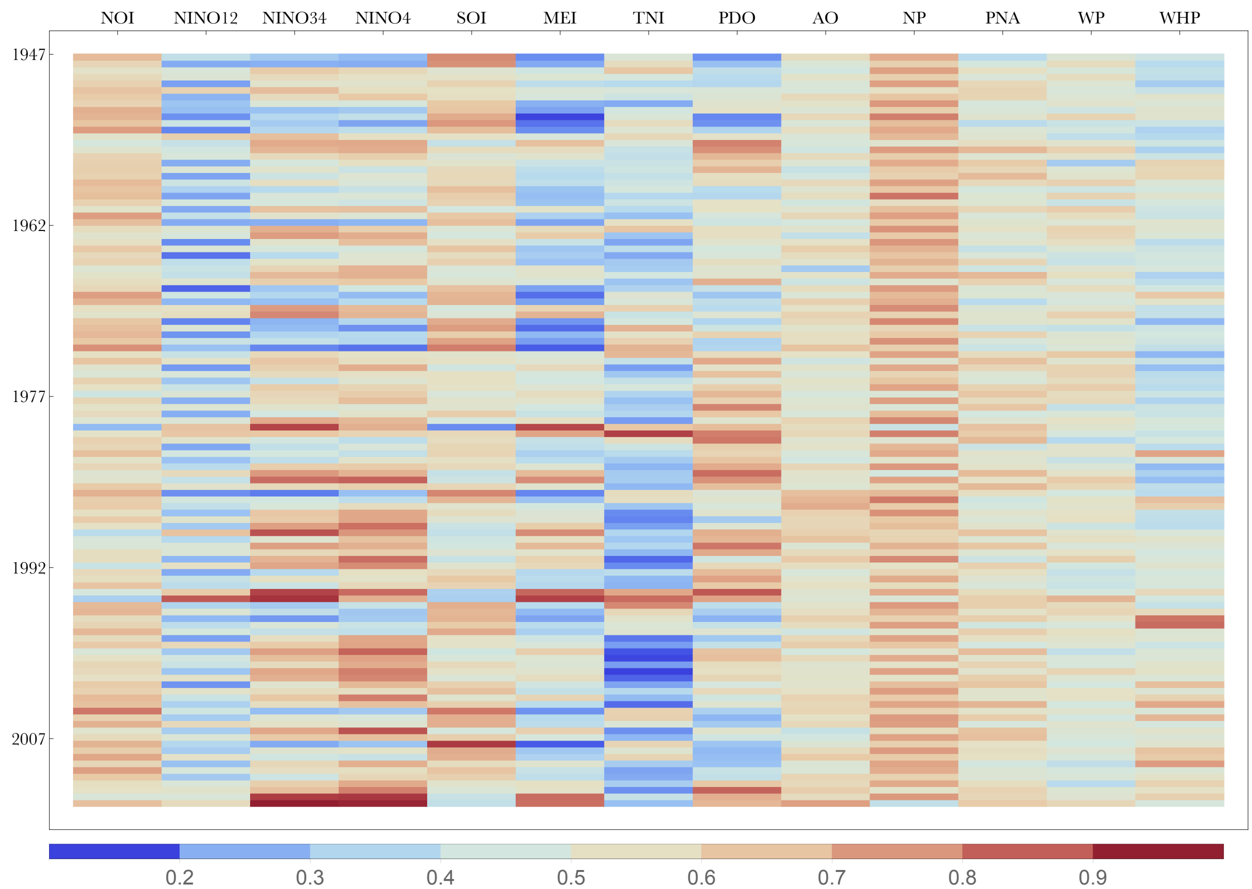


Figure 2.2: Normalized Climate Indices from 1947 to Present

Madden Julian Oscillation is another important prediction source at intraseasonal scales. Here MJO is quantified using Real-Time Multivariate MJO Index(RMM), which is composed of the first two principal components of the combined field by averaging outgoing longwave radiation, zonal wind at 850hpa and 200hpa from 15°S to 15°N[43].

Chapter 3

Methodology

3.1 Predictability Assessment

3.1.1 Evaluation Strategy

This part proposes to evaluate current dynamical models' predictability across temporal and spatial scales. Given the fact that the evaluation would be transiting form deterministic weather forecast to probabilistic climate prediction, it is important to set proper scale in evaluating forecast at different lead time. For instance, the winter weather for mid latitude is largely dominated by baroclinic disturbances, it could be expected that prediction of the 15th day's weather is not comparable to that of the first day's. Also, prediction of regional average condition should be more reliable than grid scale, due to the spatial bias in predicting cyclone generations. If too strict an acceptable threshold of model performance is put on extended forecast, we might lose the useful information at loosened spatial temporal scales, which is still of significant practical use. Given these facts, I propose the following evaluation experiments:

1. Evaluation at daily, grid scale. Evaluation is conducted on nth day grid scale precipitation forecast, n ranges from 1 to model forecast extension. For each climate division, evaluation results are spatial averaged.

2. Evaluation at daily, regional scale. For each climate division, daily regional precipitation forecasts are obtained through spatial averaging, and evaluated against CPC daily regional average records.

3. Evaluation at temporal interval, grid scale. Evaluation is conducted on grid scale precipitation averaged across certain temporal intervals. It is natural to set longer intervals as forecast lead time increases. Here we adopt the $nd nd$ strategy[44] to determine the interval windows for different lead time. Specifically, the verification is performed using data averaged over time windows equal in length to the forecast lead time. The scheme is further illustrated in Figure 3.1.

4. Evaluation at temporal interval, regional scale. Precipitation forecasts are first spatial averaged and evaluated at temporal interval scales.

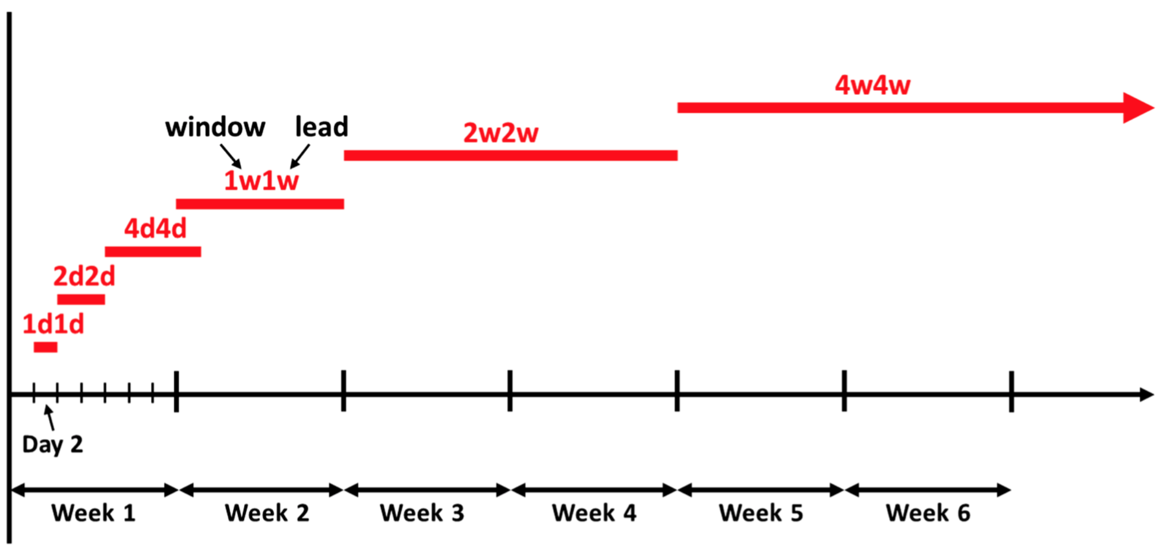


Figure 3.1: Schematic of the time window and lead time definitions used in this analysis. The horizontal axis represents forecast time from the initial condition. The expression "$1d1d$" refers to an averaging window of 1 day at a lead time of 1 day. Similarly, "$2d2d$" represents an averaging window of 2 days at a lead time of 2 days, and so on. Note that 1d1d is what is often called "day 2" in other publications, and $1w1w$ is what is often called "week 2"[44].

3.1.2 Skill Metrics

For each hindcast case, variance of ensemble members is relatively small within the medium temporal range. Deviations are assumed to be capable to be averaged out, deriving the ensemble mean as deterministic prediction. As lead time increases, ensemble members begin to diverge due to the chaotic nature of the weather system. Each member should be taken as a possible reality, although they might be distinct. For this case, probabilistic skill metrics are preferred. Since there is no definite division of determinism, to be comprehensive, we apply both deterministic skill metrics and probabilistic metric for each experiment conceived above.

Deterministic Skill Metrics

The deterministic skill measures the "distance" between ensemble mean prediction and observation. Here we adopted two common metrics, namely Pearson Correlation Coefficient(r) and Nash Sutcliffe Model Efficiency(NSE). Their formulas are listed as follows.

$$r = \frac{E[(P_{obser} - \overline{P_{obser}})(P_{simu} - \overline{P_{simu}})]}{\sigma_{P_{obser}}\sigma_{P_{simu}}}$$

$$NSE = 1 - \frac{\sum(P_{obser} - P_{simu})^2}{\sum(P_{obser} - \overline{P_{obser}})^2}$$

P_{obser} denotes precipitation observation, $\overline{P_{obser}}$ denotes mean value of precipitation observations. P_{simu} denotes ensemble mean prediction. Operator E denotes the expectation taken over all available samples, σ denotes standard deviation. r quantifies linear correlation, NSE quantifies the relative magnitude of square error compared to climatology variance.

Probabilistic Skill Metric

The Relative Operating Characteristics(ROC) score is used as probabilistic skill metric here. The reason to adopt ROC is that it provides a complete summary of hit rate and false alarm rate regarding different intervals of observation distribution, which directly relates to decision-theoretic approach and so can be easily related to the economic value of probability forecasts for forecast users[45].

To calculate ROC, the hit rates and false alarm rates for observation intervals $(-\infty, x)$(here x is set as 10 deciles of observation range) are calculated and scatter plotted(hit rate on the vertical axis and false alarm rate on the horizontal axis). The points constructs the ROC curve, which should be above 1:1 line if model held positive skill. The ROC score is defined as the area below the ROC curve. The closer the ROC score is to 1, the better. A no-skill forecast has an ROC score of 0.5[46]. Specific interpretations of ROC curve and ROC score would be illustrated with examples in the later part.

3.2 Convolutional Neural Network

3.2.1 Architecture

Basic deep learning algorithms, such as fully connected Artificial Neural Network(ANN), have long been applied for classification/regression problems in the geo community[47, 48]. While the spatial structures of inputs are very often ignored in these lumped models, they could be utilized for more compact and effective information abstraction, thereafter leading to better classification/regression results. In CNN, this is achieved through sequentially applying two peculiar matrix operators, namely convolutional layer and pooling layer.

Each convolutional layer operator filters the input field with a same kernel, which is a m×n matrix consisted of weight parameters. m and n are called kernel size.

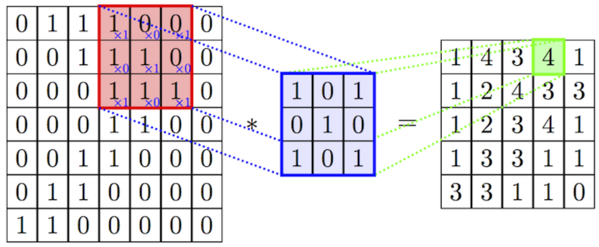


Figure 3.2: Example of convolutional layer operator[49]: the 3×3 kernel filters the left matrix by summing up corresponding elements' multiplication. Kernel moving stride is 1. Result is shown as the right matrix.

The pooling layer acts as a sub-sampling layer, which can also be viewed as a convolutional layer with a maximum value selection kernel.

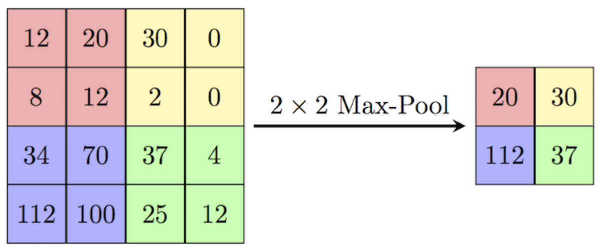


Figure 3.3: Example of pooling layer operator[49].

CNN differs from existed pattern recognition algorithms in that, instead of seeking prefixed/engineered features(or in other words, prefixed/engineered kernels), kernels of the convolutional layers could be optimized through backpropagation, just as any other parameters in the fully connected ANN. Thus, both the feature selection and following classification/regression process are trainable. Through optimized convolution and pooling operations, CNN maintains and extracts input's structures toward better representation for a given target. Besides, the sharing parameter scheme can effectively reduce overfitting by lowering model complexity. State-of-the-art regularization methods would be included, such as dropout[50] and batch normalization[51], to further balance training and test performances.

A specific example of connection precipitation distribution with circulation filed using CNN is given in Figure 3.4.

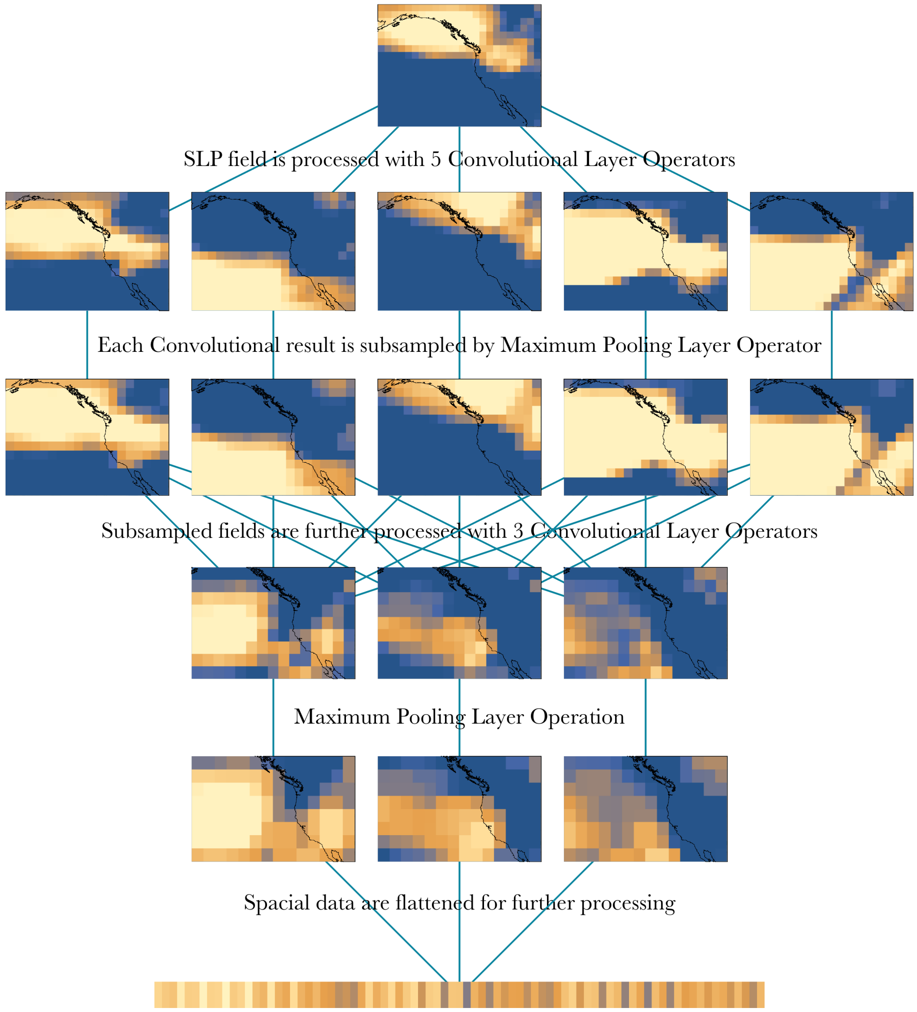


Figure 3.4: Illustration of CNN applied on the Mean Sea Level Pressure field of December, 2015. The first layer is the original Mean Sea Level Pressure(SLP) field. The second and fourth layer are convolutional layers, they use kernels to filter through their inputs to generate output. The third and fifth layer are pooling layers, they perform sub-sampling. The last layer is called flatten layer because it flattens the 2D image, it would be further processed using a fully connected neural network to regress precipitation distribution.

3.2.2 On the Analytic Property of CNN

Machine learning algorithms are usually charged as "Black Box" models despite their good estimations. For CNN, since it preserves input structures through sequential computing, we can interpret the functionality of sub components in the input field by systematically perturbing them and monitoring the change of the model output[52]. Mathematically, this is to estimate the Jacobian Matrix of the CNN, which is defined as follows:

$$J_{\text{CNN}} = \frac{\partial(Y_1, Y_2, Y_3, ..., Y_m)}{\partial(\mathrm{X}_1, \mathrm{X}_2, \mathrm{X}_3, ..., \mathrm{X}_n)} = \begin{pmatrix} \frac{\partial Y_1}{\partial \mathrm{X}_1} & \frac{\partial Y_1}{\partial \mathrm{X}_2} & \cdots & \frac{\partial Y_1}{\partial \mathrm{X}_n} \\ \frac{\partial Y_2}{\partial \mathrm{X}_1} & \frac{\partial Y_2}{\partial \mathrm{X}_2} & \cdots & \frac{\partial Y_2}{\partial \mathrm{X}_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial Y_m}{\partial \mathrm{X}_1} & \frac{\partial Y_m}{\partial \mathrm{X}_2} & \cdots & \frac{\partial Y_m}{\partial \mathrm{X}_n} \end{pmatrix}$$

Here X denotes model input(of dimension n) and Y denotes model output(of dimension m).

3.3 Regional Climate Model

The Advanced Research Weather Research and Forecasting model(WRF-ARW)[53] would be applied here for dynamical downscaling and data assimilation experiments.

The dynamical core of WRF-ARW is based on Eulerian solver for the preliminary equations. The solver uses a third-order Runge-Kutta time-integration scheme coupled with a split-explicit 2nd-order time integration scheme for the acoustic and gravity-wave modes[54].

WRF-ARW also provides options for most of the popupar physical parameterization schemes, including microphysics, cumulus, boundary layer, land surface and radiation. It offers a great flexibility in exploring model and data connections.

Chapter 4

Primary Results

4.1 Predictability Assessment

S2S Hindcasts data from 11 operational centers General Circulation Models that ran across 20 more years with restarting frequency of few days have been collected. Results of deterministic evaluation using correlation coefficient were shown in Figure 4.1.

For day by day evaluation(first two columns), as could be expected, all models show decrease of r as forecast extends. r reaches 0.2 at approximately day 10. A comparison between Column 1 and Column 2 shows that skill improvement through regional averaging is more significant for the first few days compared to extended range. There are performance differences among models. For instance, given a specific performance threshold, forecast extension between models differ by up to 4 days approximately.

Smoother transition of model skills across forecast lead time is achieved by evaluating models at temporal interval scales(third and forth column). r for the first week is above the order of 0.6 at grid scale and 0.7 at regional scale. JMA, KMA and ECCC models lead the best performance at this temporal range. r for the second week($1w1w$) is of the order of 0.5 at grid scale and 0.6 at regional scale, considering the best case(ECMWF). The average performance is of the order of 0.4 for both grid and regional scale. It is noteworthy that some models show unexpected good performance for longer forecast ranges(beyond two weeks), such as BoM for Southern California and HMCR for Washington State.

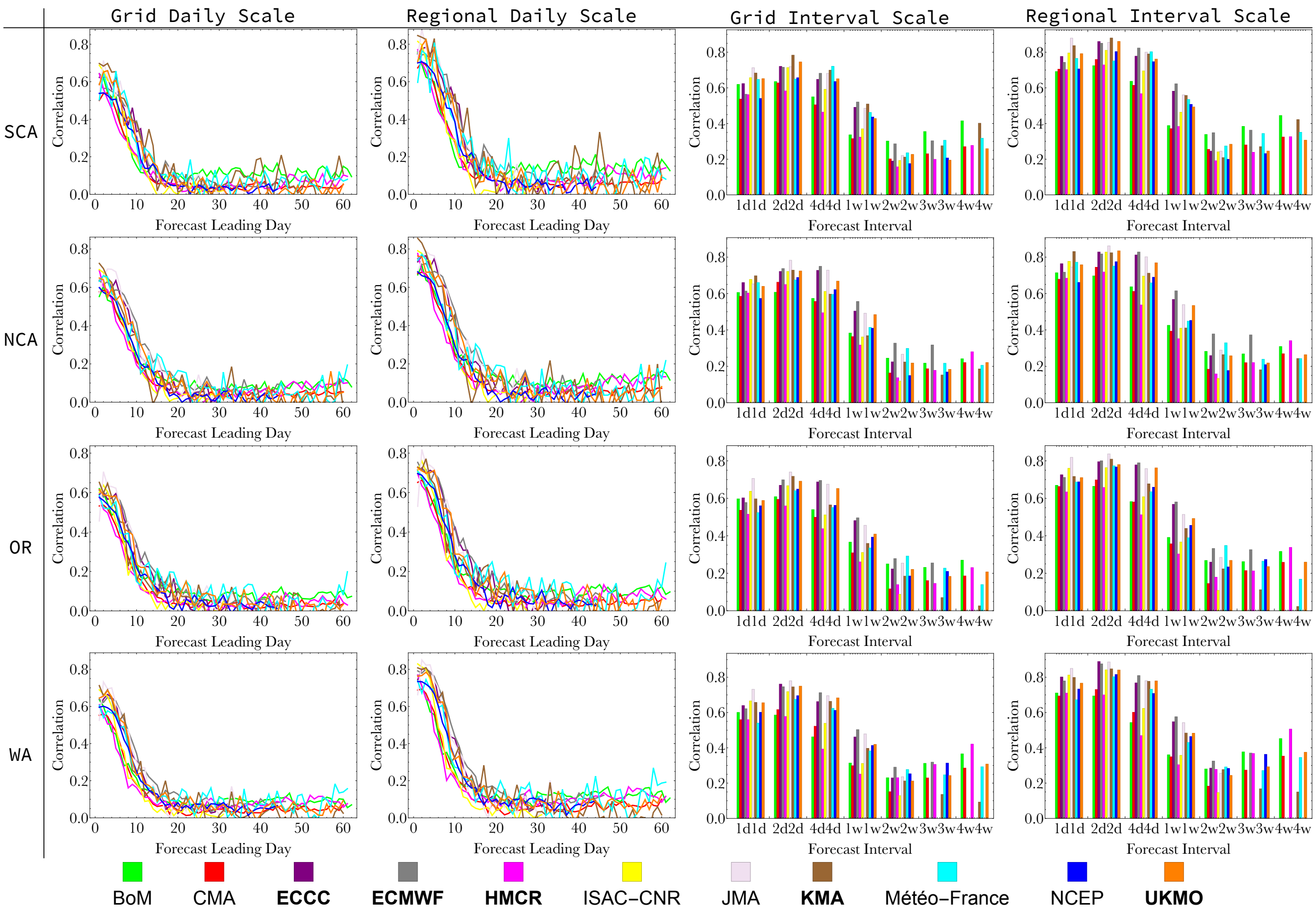


Figure 4.1: Correlation coefficient between precipitation ensemble mean prediction and observations for 4 experiments in 4 climate divisions.

4.2 Connecting Precipitation to Circulation Using Convolutional Neural Network

In order to take full use of long term dynamical forecast for extended precipitation prediction, I applied CNN to seek the existence of less-resolved but more stable connection between precipitation and fundamental climate variables.

The experiments were done through averaging Sea Level Pressure(SLP) and precipitation field data(from reanlaysis and observation) to different temporal scales and establish their connections using CNN explained in the previous section.

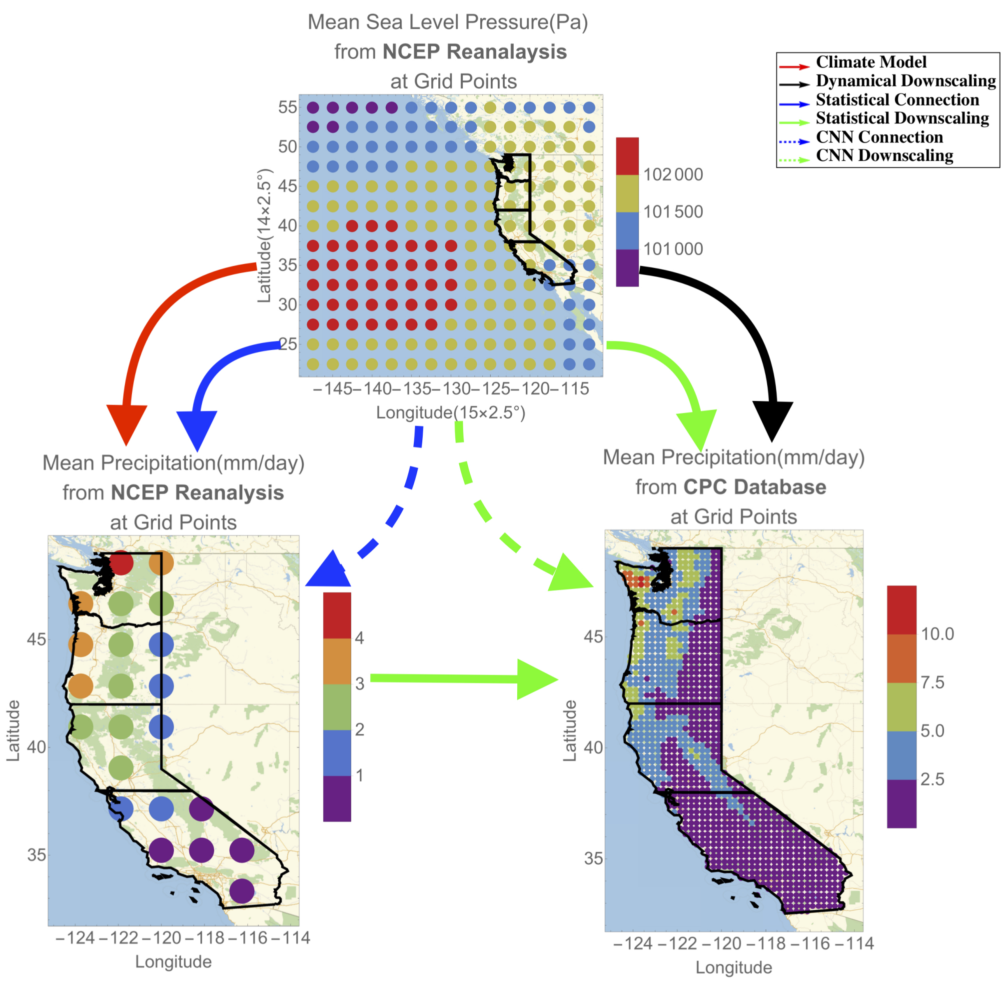


Figure 4.2: Connections between SLP and precipitation using different models

Precipitation estimation results(compared against reanalysis in Figure 4.1 and against observation in Figure 4.2) at different scales were shown as follows:

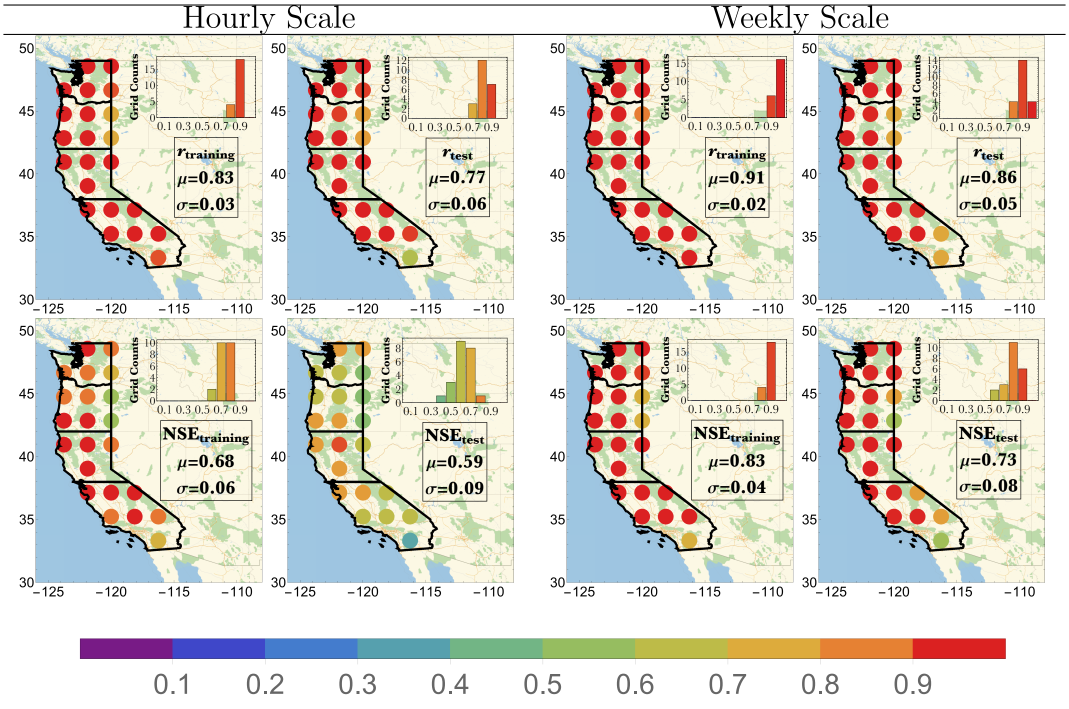


Table 4.1: Precipitation estimation results compared against reanalysis at hourly and weekly scales

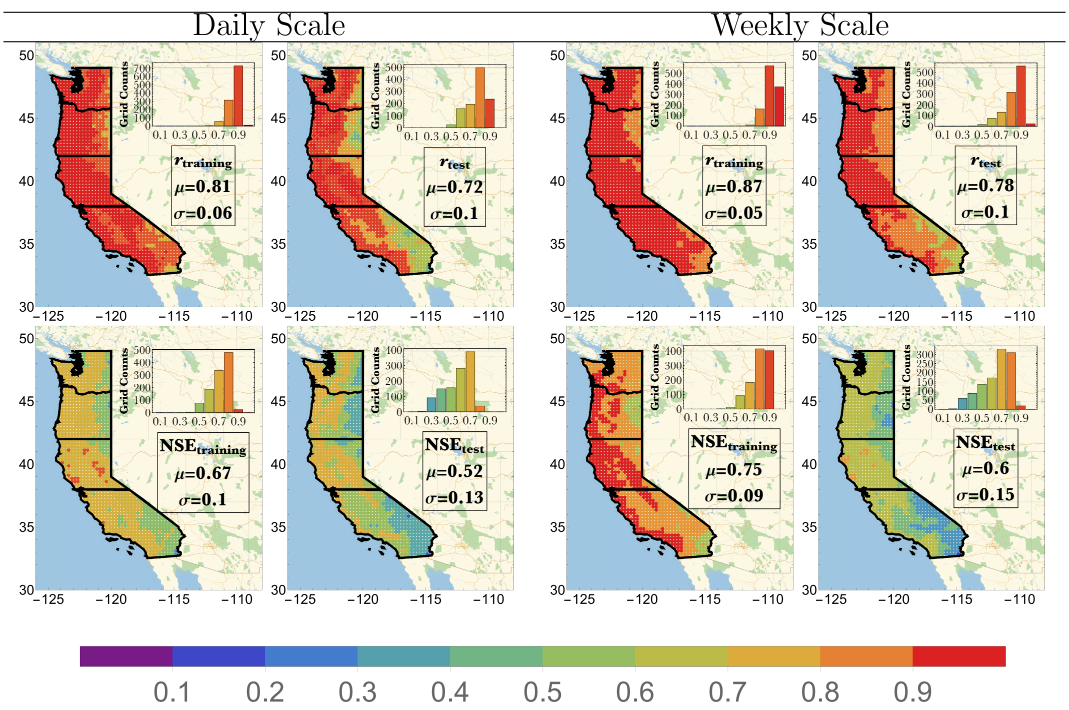


Table 4.2: Precipitation estimation results compared against observation at daily and weekly scales

A summary of CNN performance against two baseline models, namely Random Forest and Linear Regression is given below.

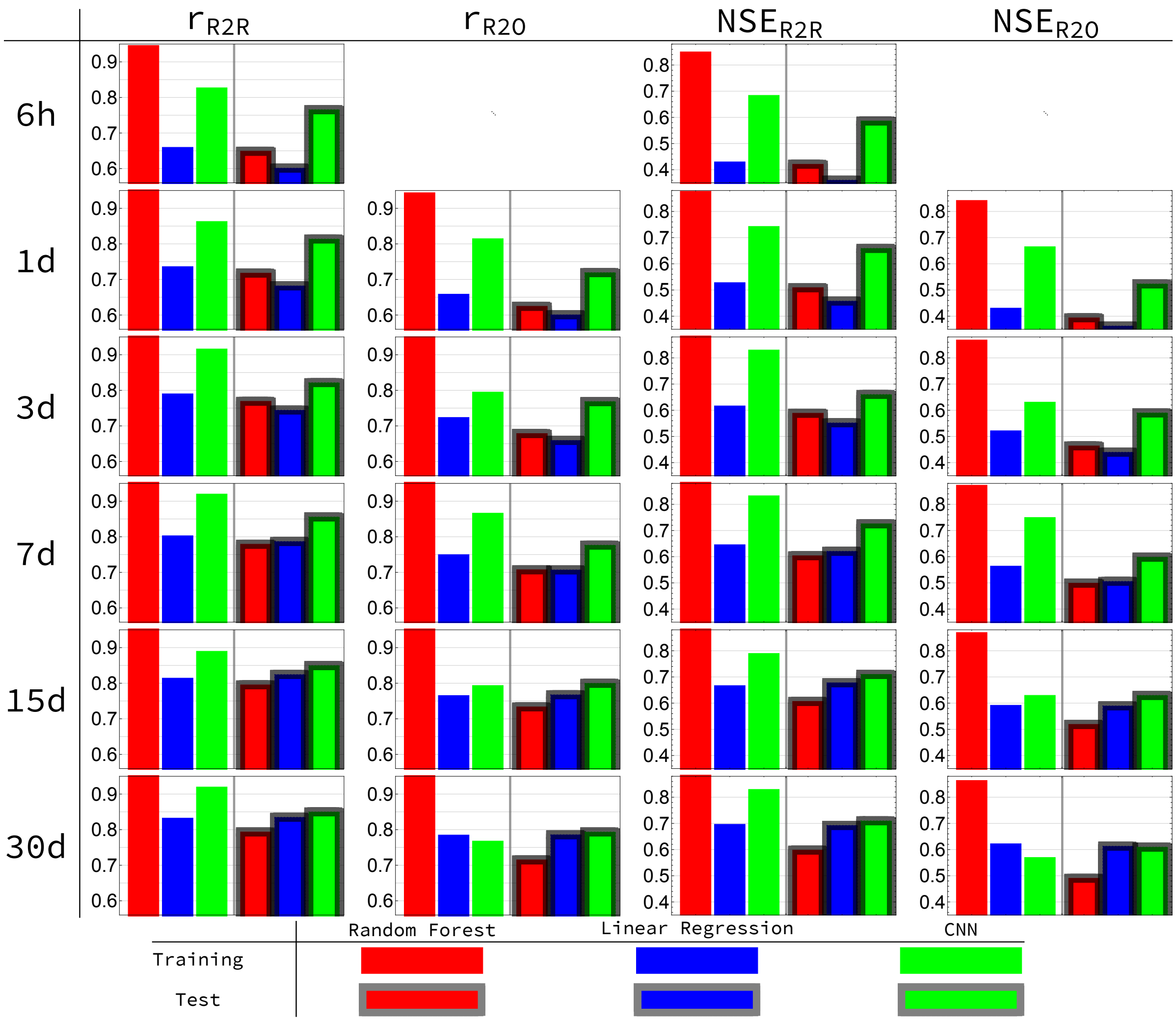


Figure 4.3: Summary of model performance at different temporal scales.

CNN showed significant advantage over baseline models at hourly to weekly scales. Random Forest model showed consistent overfitting(test performance being much poorer than training performance) across scales. In fact, this is quite common for classical machine learning algorithms when applied for image or higher dimensional data problems. Performances at training and test set for Linear Regression are in accordant and increase simultaneously with temporal scales, reaching similar level to CNN at 15d scale.

Previous work using linear connection between SLP and P showed that the connection can provide equally good precipitation prediction using numerical models' SLP, compared to original model precipitation estimation results[55]. Since CNN applied here outperforms linear connection for certain scales, it is quite promising to produce better precipitation estimations.

4.3 Influence of ENSO on Predictability in West Coast United States

As the major intraannual climate variability, ENSO exerts direct influences on tropics and impacts extratropics through its teleconnections. Although no consensus has been reached about its specific influence on precipitation at intraseasonal to intraaanual scales for the study area of West Coast United States, it is reasonable to assume that the circulation dynamics should be perturbated by the ENSO phases, and the predictable time could be different, which opens up predictability opportunities for the area.

To explore ENSO as potential prediction windows, I first clustered hindcast cases according to their launching time's ENSO phase, defined here using the Niño 34 index. The abundance of hindcast cases in the S2S database guarantees that for each phase there are enough samples to reach convincing conclusions.

Considering Week 2 prediction, as shown in Figure 4.4, for Southern California, all models show better deterministic performances for El Niño phase. The advantage of El Niño phase prediction over La niña is more significant in better performing models, such as for ECCC, ECMWF,KMA and UKMO. For ECMWF in El Niño, the regional prediction correlation reaches 0.7, compared to 0.6 for indiscriminate evaluation. We also noticed that although ECMWF performs best at Week 2 in general, KMA model shows better performance in ENSO neural phase for the area, considering all the skill metrics.

For Northern California, 8 out of 11 models show better deterministic skills in La Niña phase. Model performance difference is not as significant as for cases in SCA, except the

JMA model. We noticed that currently JMA only provides winter hindcasts of October to December. Further exploration should be conducted to determine whether the distinction for ENSO phases is due to model or climatology.

For Oregon, 10 out of 11 models show better performance in La Niña phase, except MeteoFrance model, and it is actually doing pretty good for El Niño and neutral phases. Again we found that the advantage of La niña phase prediction over El Niño is more significant in better performing models, such as for ECCC, ECMWF,JMA,KMA and UKMO.

For Washington State, models did not show consistent predictability difference considering ENSO phases. However, we noticed that MeteoFrance model keeps showing significant better performance for El Niño and neutral phases compared to La Niña, same situation for Oregon.

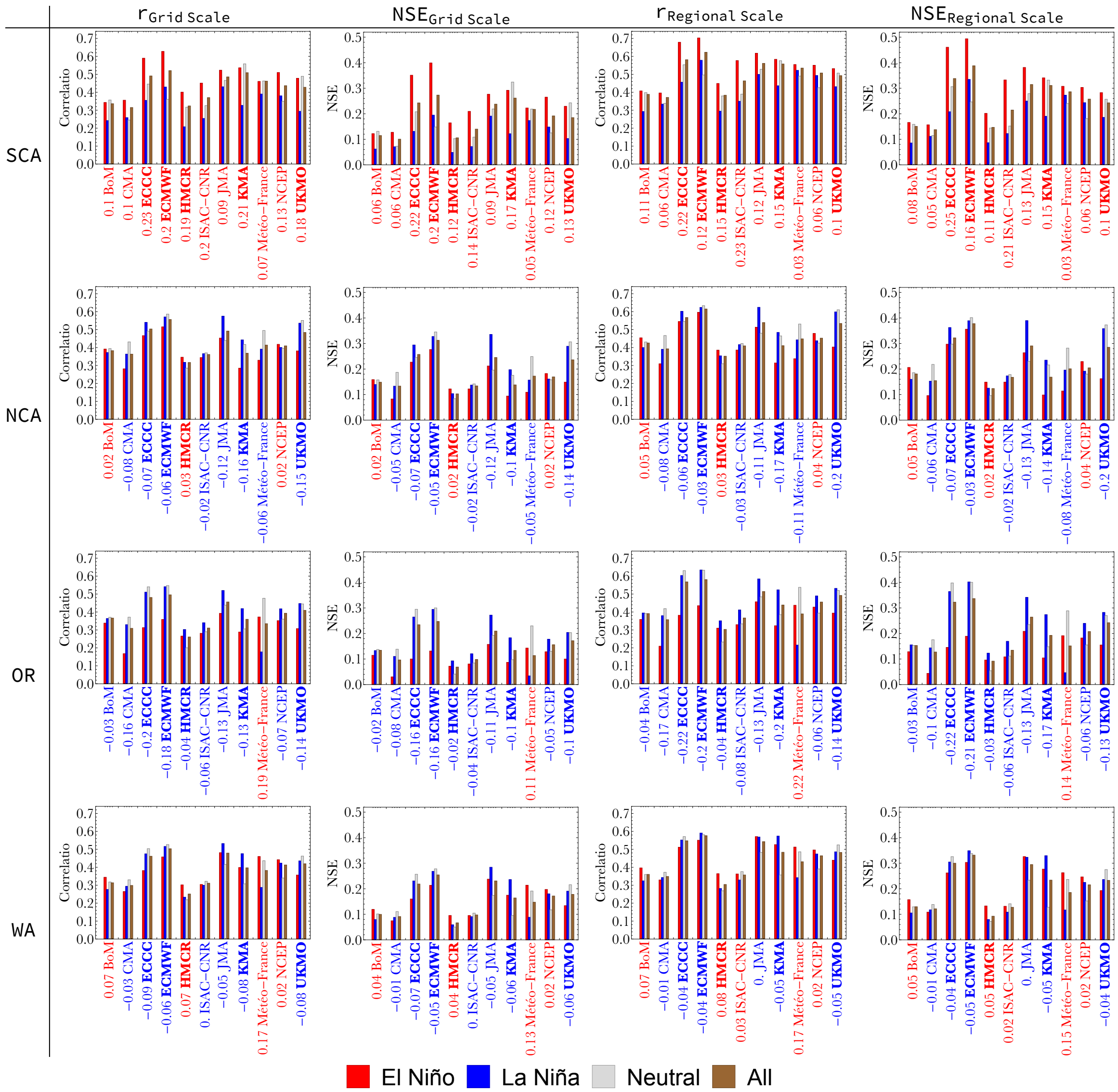


Figure 4.4: Week 2 deterministic predictability for different ENSO phases. Red bar denotes skill for El Niño phase, blue for La Niña, grey for neutral, brown for all time. Model is labeled red if performance for El Niño phase is better than La Niña, vice versa. Number denotes the skill difference between El Niño and La Niña.

Considering Week 3∼4 prediction, as shown in Figure 4.5, for Southern California, still there are around 9 out of 11 models showing better deterministic skills in El Niño phase, especially for better performing models.

For Northern California, ECMWF model shows salient performance in El Niño phase at regional scale($r = 0.45$ and $NSE = 0.21$). MeteoFrance model shows salient performance in Neutral phase($r = 0.41$ and $NSE = 0.17$).

For Oregon and Washington State, most models show better deterministic skills in La Niña phase. An exception is the KMA model, which shows particular advance in El Niño phase.

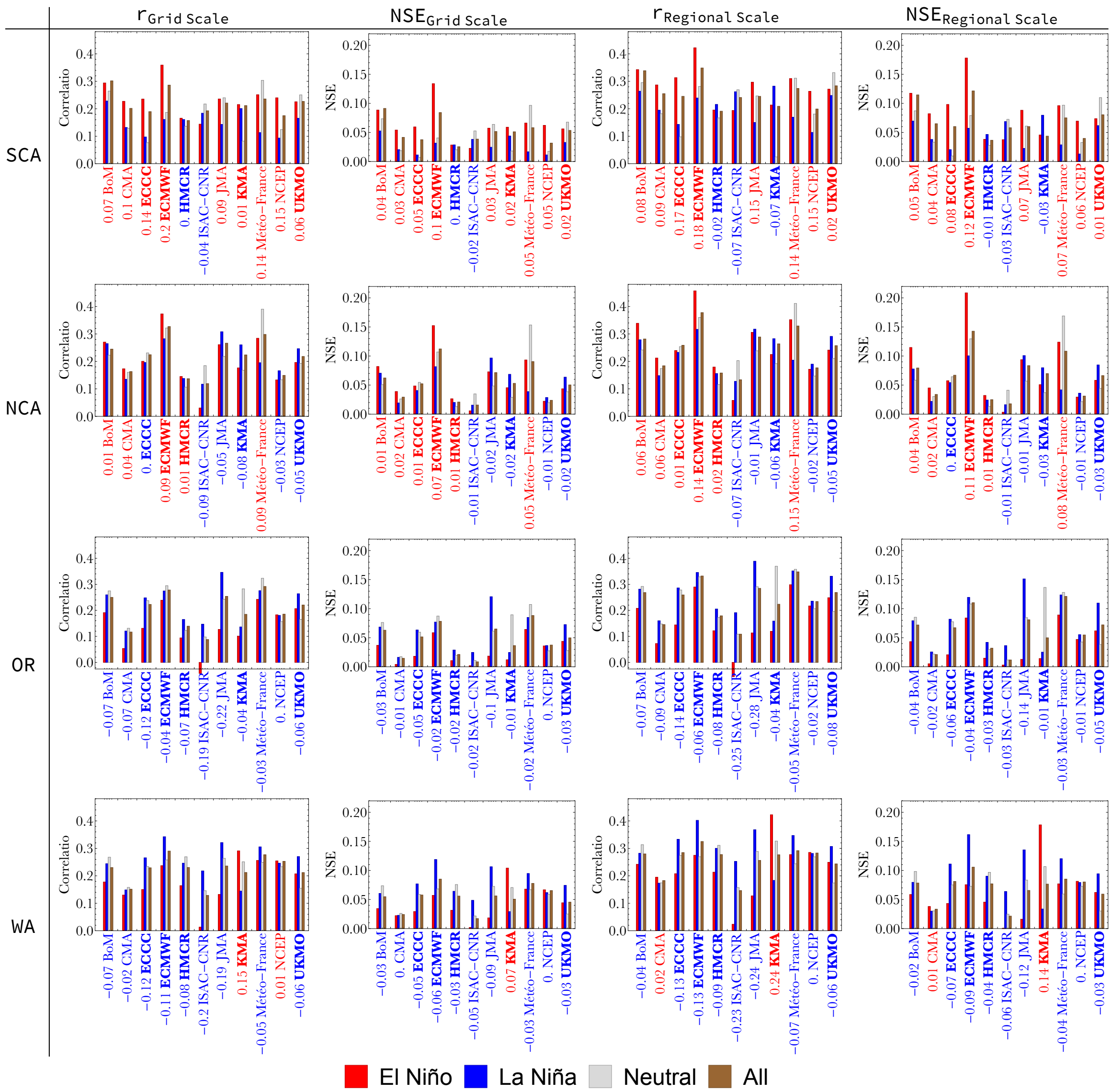


Figure 4.5: Week 3∼4 deterministic predictability for different ENSO phases. Same denotation as in Figure 4.4.

4.4 WRF Simulation of Extratropical Cyclone Events

The WRF-ARW model has been installed and tested on the Hyper Performance Computing(HPC) nodes. Below I show a case study of extratropical cyclone generation case for West Coast United States on February 10-20, 2010.

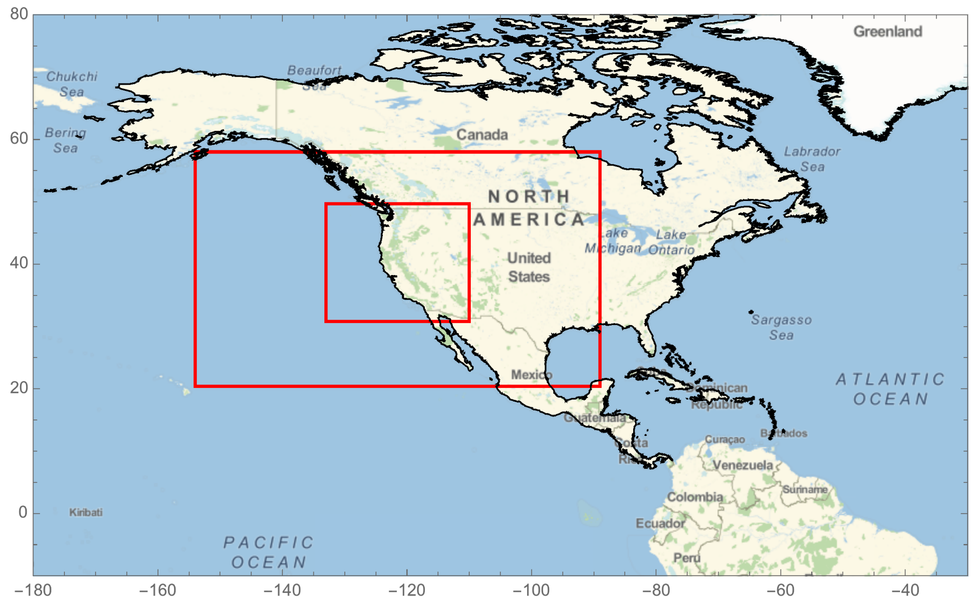


Figure 4.6: WRF domains

The simulated precipitation process was shown as follows.

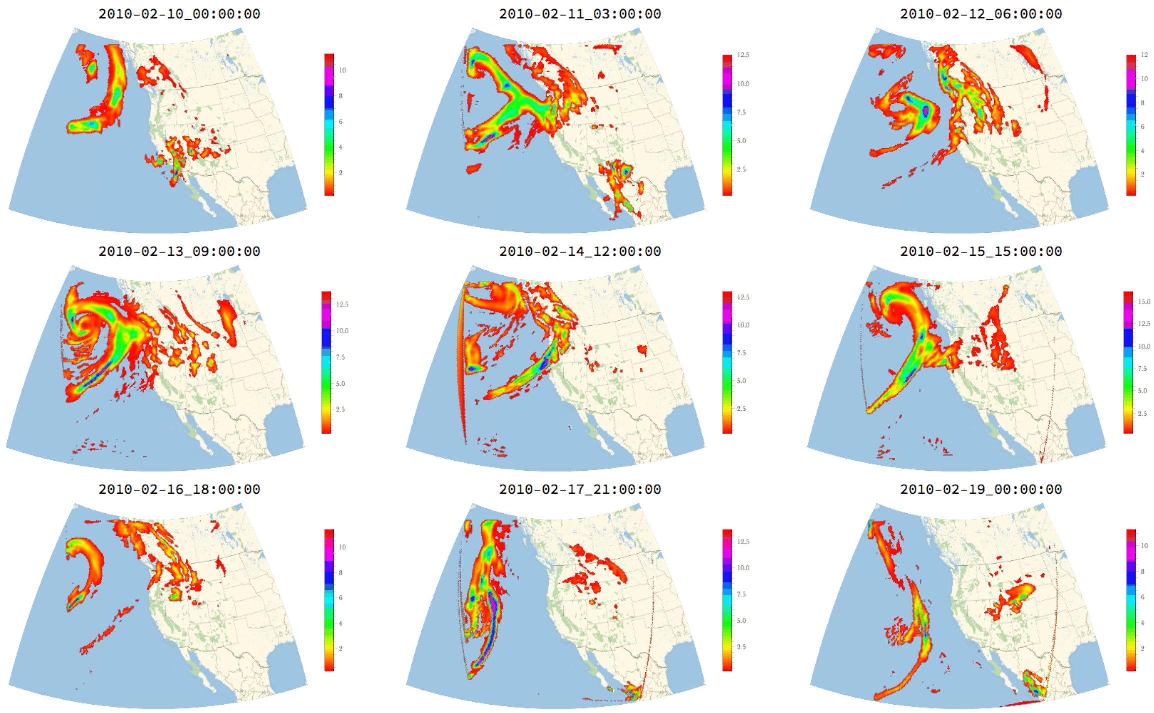


Figure 4.7: WRF simulation of precipitation process

Chapter 5

Future Work and Schedules

The following works are scheduled to be finished in the thesis work.

1. Make a more comprehensive evaluation on predictability status quo of general circulation models. Existed evaluation work focus on total precipitation amount only, using limited deterministic skills and no probabilistic skills. Evaluation is also restricted to the West Coast United States. The future work would apply both deterministic and probabilistic skills for more features(mean, extreme, etc.) of precipitation and temperature covering a larger spacial scope.

2. Further decrease precipitation parameterization errors by exploring the possibility of estimating precipitation from resolved, more accurately estimated dynamical variables across different scales.

   (a) Explore the next generation of image processing algorithm, namely, Capsule Neural Network(CapsNet) for a better estimation of precipitation circulation connection. As an improvement to CNN that achieves state-of-the-art performance in testing dataset, CapsNet preserves detailed information of objects location and its pose, which might contribute to even better results.

   (b) Simple temporal averaging might lose useful information as pressure features(depression, for instance) might be averaged out. Try the convolutional recurrent neural network structure to distill more information for precipitation estimation.

(c) Use the constructed connection to process forecast SLP for precipitation estimation, and compare to original model precipitation products.

(d) Test if the circulation-precipitation connection holds true for different regions, especially for areas that hold limited moisture for convection. Try to incorporate water constrain information in the framework to make it more universal.

3. Verify the significance of ENSO's influence on predictability. Also, test if MJO and other potential prediction sources might influence prediction performances. If the assumption that climate scenario defined by climate indices could leave significant influence on model predictability, these hindcast empirical knowledge could help to build a multi model ensemble strategy based on model performance at different scenarios.

4. Figure out the optimum resolution, parameterization settings for running WRF-ARW. Inspect the connection between precipitation and dynamical variables using its high resolution outputs. Develop flood nowcast taking use of both numerical models and remote sensing data within the framework of recurrent convolutional network.

Schedule for this research work is shown as follows:

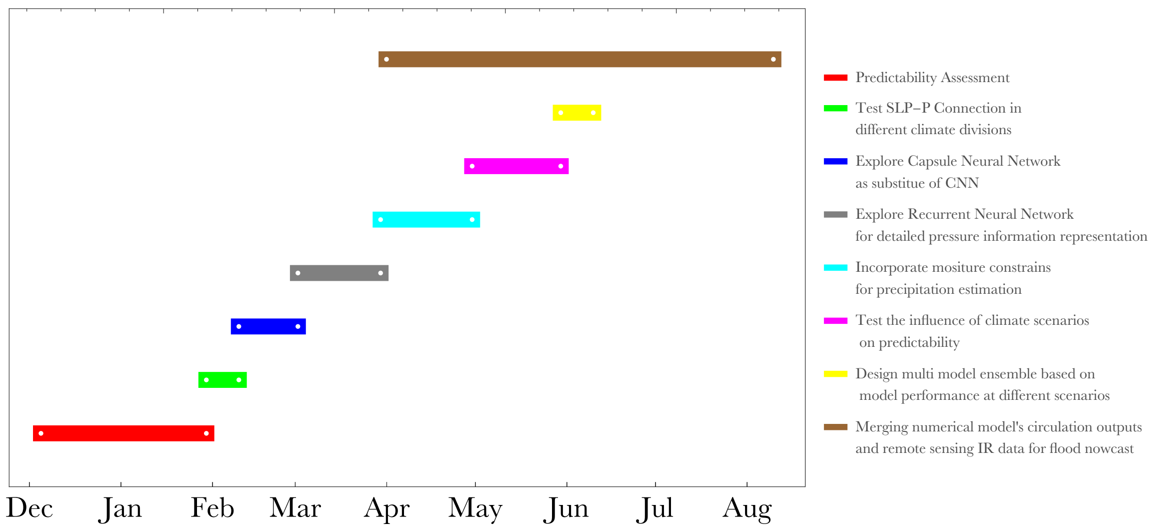


Figure 5.1: Schedule

Bibliography

[1] Andrew W Robertson, Arun Kumar, Malaquias Peña, and Frederic Vitart. Improving and promoting subseasonal to seasonal prediction. *Bulletin of the American Meteorological Society*, 96(3):ES49–ES53, 2015.

[2] Christopher J White, Henrik Carlsen, Andrew W Robertson, Richard JT Klein, Jeffrey K Lazo, Arun Kumar, Frederic Vitart, Erin Coughlan de Perez, Andrea J Ray, Virginia Murray, et al. Potential applications of subseasonal-to-seasonal (s2s) predictions. *Meteorological Applications*, 2017.

[3] Vilhelm Frieman Koren Bjerknes. Fields of force. 1906.

[4] Adrian J Simmons and Anthony Hollingsworth. Some aspects of the improvement in skill of numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 128(580):647–677, 2002.

[5] Peter Bauer, Alan Thorpe, and Gilbert Brunet. The quiet revolution of numerical weather prediction. *Nature*, 525(7567):47–55, 2015.

[6] Shuhua Li and Andrew W Robertson. Evaluation of submonthly precipitation forecast skill from global ensemble prediction systems. *Monthly Weather Review*, 143(7):2871–2889, 2015.

[7] Frédéric Vitart. Evolution of ecmwf sub-seasonal forecast skill scores. *Quarterly Journal of the Royal Meteorological Society*, 140(683):1889–1899, 2014.

[8] Eugenia Kalnay, Stephen J Lord, and Ronald D McPherson. Maturity of operational numerical weather prediction: Medium range. *Bulletin of the American Meteorological Society*, 79(12):2753–2769, 1998.

[9] Peter Lynch. The origins of computer weather prediction and climate modeling. *Journal of Computational Physics*, 227(7):3431–3444, 2008.

[10] David J Stensrud. *Parameterization schemes: keys to understanding numerical weather prediction models*. Cambridge University Press, 2009.

[11] Wikipedia. Plagiarism — Wikipedia, the free encyclopedia, 2017. [Online; accessed 22-July-2017].

[12] Erica Lucy Thompson. *Modelling North Atlantic storms in a changing climate*. PhD thesis, Imperial College London, 2013.

[13] Edward N Lorenz. A study of the predictability of a 28-variable atmospheric model. *Tellus*, 17(3):321–333, 1965.

[14] Charles Jones, Duane E Waliser, KM Lau, and W Stern. The madden–julian oscillation and its impact on northern hemisphere weather predictability. *Monthly weather review*, 132(6):1462–1471, 2004.

[15] Ocean Studies Board, Engineering National Academies of Sciences, Medicine, et al. *Next generation earth system prediction: strategies for subseasonal to seasonal forecasts*. National Academies Press, 2016.

[16] Michael Ghil and Andrew W Robertson. Solving problems with gcms: General circulation models and their role in the climate modeling hierarchy. *International Geophysics Series*, 70:285–326, 2000.

[17] M Ghil. Hilbert problems for the geosciences in the 21st century. *Nonlinear processes in geophysics*, 8(4/5):211–211, 2001.

[18] Kuo-lin Hsu, Hoshin Vijai Gupta, and Soroosh Sorooshian. Artificial neural network modeling of the rainfall-runoff process. *Water resources research*, 31(10):2517–2530, 1995.

[19] Soroosh Sorooshian, Kuo-Lin Hsu, Xiaogang Gao, Hoshin V Gupta, Bisher Imam, and Dan Braithwaite. Evaluation of persiann system satellite–based estimates of tropical rainfall. *Bulletin of the American Meteorological Society*, 81(9):2035–2046, 2000.

[20] David Silverman and John A Dracup. Artificial neural networks and long-range precipitation prediction in california. *Journal of applied meteorology*, 39(1):57–66, 2000.

[21] Why deep learning. `https://www.slideshare.net/ExtractConf`. Accessed: 2017-09-30.

[22] Benjamin Klein, Lior Wolf, and Yehuda Afek. A dynamic convolutional layer for short range weather prediction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4840–4848, 2015.

[23] SHI Xingjian, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo. Convolutional lstm network: A machine learning approach for precipitation nowcasting. In *Advances in Neural Information Processing Systems*, pages 802–810, 2015.

[24] Yunjie Liu, Evan Racah, Joaquin Correa, Amir Khosrowshahi, David Lavers, Kenneth Kunkel, Michael Wehner, William Collins, et al. Application of deep convolutional neural networks for detecting extreme weather in climate datasets. *arXiv preprint arXiv:1605.01156*, 2016.

[25] Evan Racah, Christopher Beckham, Tegan Maharaj, Christopher Pal, et al. Semi-supervised detection of extreme weather events in large climate datasets. *arXiv preprint arXiv:1612.02095*, 2016.

[26] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[27] Pingping Xie, Mingyue Chen, Song Yang, Akiyo Yatagai, Tadahiro Hayasaka, Yoshihiro Fukushima, and Changming Liu. A gauge-based analysis of daily precipitation over east asia. *Journal of Hydrometeorology*, 8(3):607–626, 2007.

[28] SJ Brown, J Caesar, and Christopher AT Ferro. Global changes in extreme daily temperature since 1950. *Journal of Geophysical Research: Atmospheres*, 113(D5), 2008.

[29] Eugenia Kalnay, Masao Kanamitsu, Robert Kistler, William Collins, Dennis Deaven, Lev Gandin, Mark Iredell, Suranjana Saha, Glenn White, John Woollen, et al. The ncep/ncar 40-year reanalysis project. *Bulletin of the American meteorological Society*, 77(3):437–471, 1996.

[30] Sakari M Uppala, PW Kållberg, AJ Simmons, U Andrae, V d Bechtold, M Fiorino, JK Gibson, J Haseler, A Hernandez, GA Kelly, et al. The era-40 re-analysis. *Quarterly Journal of the royal meteorological society*, 131(612):2961–3012, 2005.

[31] F Vitart, C Ardilouze, A Bonet, A Brookshaw, M Chen, C Codorean, M Déqué, L Ferranti, E Fucile, M Fuentes, et al. The subseasonal to seasonal (s2s) prediction project database. *Bulletin of the American Meteorological Society*, 98(1):163–173, 2017.

[32] Oscar Alves, Guomin Wang, Aihong Zhong, Neville Smith, F Tseitkin, Graham Warren, Andreas Schiller, Stuart Godfrey, and Gary Meyers. Poama: Bureau of meteorology operational coupled model seasonal forecast system. In *Proceedings of national drought forum, Brisbane*, pages 49–56, 2003.

[33] Tongwen Wu, Lianchun Song, Weiping Li, Zaizhi Wang, Hua Zhang, Xiaoge Xin, Yanwu Zhang, Li Zhang, Jianglong Li, Fanghua Wu, et al. An overview of bcc climate system model development and application for climate change studies. *Journal of Meteorological Research*, 28(1):34–56, 2014.

[34] Frederic Vitart, Roberto Buizza, Magdalena Alonso Balmaseda, Gianpaolo Balsamo, Jean-Raymond Bidlot, Axel Bonet, Manuel Fuentes, Alfred Hofstadler, Franco Molteni, and Tim N Palmer. The new vareps-monthly forecasting system: A first step towards seamless prediction. *Quarterly Journal of the Royal Meteorological Society*, 134(636):1789–1799, 2008.

[35] N Gagnon, XX Deng, PL Houtekamer, M Charron, A Erfani, S Beauregard, B Archambault, F Petrucci, and A Giguère. Improvements to the global ensemble prediction system (geps) from version 2.0. 3 to version 3.0. 0. *Tech. re., Development and Operations Divisions, Meteorological Research Division, Canadian Meteorological Center, Environment Canada*, 2013.

[36] Piero Malguzzi, Andrea Buzzi, and Oxana Drofa. The meteorological global model globo at the isac-cnr of italy assessment of 1.5 yr of experimental use for medium-range weather forecasts. *Weather and Forecasting*, 26(6):1045–1055, 2011.

[37] Philippe Courtier and Jean-Francois Geleyn. A global numerical weather prediction model with variable resolution: Application to the shallow-water equations. *Quarterly Journal of the Royal Meteorological Society*, 114(483):1321–1346, 1988.

[38] Japan Meteorological Agency. Outline of the operational numerical weather prediction at the japan meteorological agency, 2013.

[39] MJ Best, M Pryor, DB Clark, GG Rooney, R Essery, CB Ménard, JM Edwards, MA Hendry, A Porson, N Gedney, et al. The joint uk land environment simulator (jules), model description–part 1: energy and water fluxes. *Geoscientific Model Development*, 4(3):677–699, 2011.

[40] Aurore Voldoire, E Sanchez-Gomez, D Salas y Mélia, B Decharme, Christophe Cassou, S Sénési, Sophie Valcke, I Beau, A Alias, M Chevallier, et al. The cnrm-cm5. 1 global

climate model: description and basic evaluation. *Climate Dynamics*, 40(9-10):2091–2121, 2013.

[41] Suranjana Saha, Shrinivas Moorthi, Xingren Wu, Jiande Wang, Sudhir Nadiga, Patrick Tripp, David Behringer, Yu-Tai Hou, Hui-ya Chuang, Mark Iredell, et al. The ncep climate forecast system version 2. *Journal of Climate*, 27(6):2185–2208, 2014.

[42] Nigel Wood, Andrew Staniforth, Andy White, Thomas Allen, Michail Diamantakis, Markus Gross, Thomas Melvin, Chris Smith, Simon Vosper, Mohamed Zerroukat, et al. An inherently mass-conserving semi-implicit semi-lagrangian discretization of the deep-atmosphere global non-hydrostatic equations. *Quarterly Journal of the Royal Meteorological Society*, 140(682):1505–1520, 2014.

[43] Matthew C Wheeler and Harry H Hendon. An all-season real-time multivariate mjo index: Development of an index for monitoring and prediction. *Monthly Weather Review*, 132(8):1917–1932, 2004.

[44] Hongyan Zhu, Matthew C Wheeler, Adam H Sobel, and Debra Hudson. Seamless precipitation prediction skill in the tropics and extratropics from a global model. *Monthly Weather Review*, 142(4):1556–1569, 2014.

[45] Ian T Jolliffe and David B Stephenson. *Forecast verification: a practitioner's guide in atmospheric science*. John Wiley & Sons, 2003.

[46] Frédéric Vitart. Monthly forecasting at ecmwf. *Monthly Weather Review*, 132(12):2761–2779, 2004.

[47] Donna M Rizzo and David E Dougherty. Characterization of aquifer properties using artificial neural networks: Neural kriging. *Water Resources Research*, 30(2):483–497, 1994.

[48] Matt W Gardner and SR Dorling. Artificial neural networks (the multilayer perceptron)a review of applications in the atmospheric sciences. *Atmospheric environment*, 32(14):2627–2636, 1998.

[49] Deep learning for complete beginners: convolutional neural networks with keras. https://cambridgespark.com/content/tutorials/convolutional-neural-networks-with-keras/index.html. Accessed: 2017-04-14.

[50] Nitish Srivastava, Geoffrey E Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(1):1929–1958, 2014.

[51] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.

[52] Matthew D Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In *European conference on computer vision*, pages 818–833. Springer, 2014.

[53] J Michalakes, J Dudhia, D Gill, T Henderson, J Klemp, W Skamarock, and W Wang. The weather research and forecast model: software architecture and performance. In *Proceedings of the Eleventh ECMWF Workshop on the Use of High Performance Computing in Meteorology*, pages 156–168. World Scientific: Singapore, 2005.

[54] William C Skamarock, Joseph B Klemp, Jimy Dudhia, David O Gill, Dale M Barker, Wei Wang, and Jordan G Powers. A description of the advanced research wrf version 2. Technical report, National Center For Atmospheric Research Boulder Co Mesoscale and Microscale Meteorology Div, 2005.

[55] Stephen P Charles, Bryson C Bates, Ian N Smith, and James P Hughes. Statistical downscaling of daily precipitation from observed and modelled atmospheric fields. *Hydrological Processes*, 18(8):1373–1394, 2004.