Machine learning based algorithms for uncertainty quantification in numerical weather prediction models

Azam Moosavi[a], Vishwas Rao[b], Adrian Sandu[a]

[a] *Computational Science Laboratory, Department of Computer Science*
Virginia Polytechnic Institute and State University, Blacksburg, VA 24060, USA
E-mail: {azmosavi, asandu7}@vt.edu
[b] *Mathematics and Computer Science Division, Argonne National Laboratory, Lemont, IL, USA*
E-mail: vhebbur@anl.gov

Abstract

Complex numerical weather prediction models incorporate a variety of physical processes, each described by multiple alternative physical schemes with specific parameters. The selection of the physical schemes and the choice of the corresponding physical parameters during model configuration can significantly impact the accuracy of model forecasts. There is no combination of physical schemes that works best for all times, at all locations, and under all conditions. It is therefore of considerable interest to understand the interplay between the choice of physics and the accuracy of the resulting forecasts under different conditions.

This paper demonstrates the use of machine learning techniques to study the uncertainty in numerical weather prediction models due to the interaction of multiple physical processes. The first problem addressed herein is the estimation of systematic model errors in output quantities of interest at future times, and the use of this information to improve the model forecasts. The second problem considered is the identification of those specific physical processes that contribute most to the forecast uncertainty in the quantity of interest under specified meteorological conditions. In order to address these questions we employ two machine learning approaches, random forests and artificial neural networks. The discrepancies between model results and observations at past times are used to learn the relationships between the choice of physical processes and the resulting forecast errors.

Numerical experiments are carried out with the Weather Research and Forecasting (WRF) model. The output quantity of interest is the model precipitation, a variable that is both extremely important and very challenging to forecast. The physical processes under consideration include various microphysics schemes, cumulus parameterizations, short wave, and long wave radiation schemes. The experiments demonstrate the strong potential of machine learning approaches to aid the study of model errors.

Keywords: Numerical weather prediction model, precipitation prediction,

1. Introduction

Computer simulation models of the physical world, such as numerical weather prediction (NWP) models, are imperfect and can only approximate the complex evolution of physical reality. Some of the errors are due to the uncertainty in the initial and boundary conditions, forcings, and model parameter values. Other errors, called structural model errors, are due to our incomplete knowledge about the true physical processes, and manifest themselves as missing dynamics in the model [38]. Examples of structural errors include the misrepresentation of sea-ice in the spring and fall, errors affecting the stratosphere above polar regions in winter [58], as well as errors due to the interactions among (approximately-represented) physical processes.

Data assimilation improves model forecasts by fusing information from both model outputs and observations of the physical world in a coherent statistical estimation framework [1, 30, 43, 58]. While traditional data assimilation reduces the uncertainty in the model state and model parameter values, no methodologies to reduce the structural model uncertainty are available to date.

In this study we consider the Weather Research and Forecasting (WRF) model [61], a mesoscale atmospheric modeling system. The WRF model includes multiple physical processes and parametrization schemes, and choosing different model options can lead to significant variability in the model predictions [12, 42].

Among different atmospheric phenomena, the prediction of precipitation is extremely challenging and is obtained by solving the atmospheric dynamic and thermodynamic equations [42]. Model forecasts of precipitation are very sensitive to physics options such as the micro-physics, cumulus, long wave, and short wave radiation [13, 42, 34]. Other physics settings that can affect the WRF precipitation predictions include surface physics, planetary boundary layer (PBL), land-surface (LS) parameterizations, and lateral boundary condition. Selecting the right physical process representations and parameterizations is a challenge. In practice the values of physical parameters are empirically determined such as to minimize the difference between the measurements and model predictions [61, 34].

Considerable effort has been dedicated to determining the best physical configurations of the weather forecast models such as to improve their predictions of precipitation. No single choice of physical parameters works perfectly for all times, geographical locations, or meteorological conditions [16, 60]. Lowrey and Yang [34] investigated the errors in precipitation predictions caused by different parameters including micro-physics and cumulus physics, the buffer zone, the initialization interval, the domain size, and the initial and boundary conditions. Jankov et al. [26] examined different combinations of cumulus convection schemes, micro-physical options, and boundary conditions. They concluded that no configuration was the clear winner at all times, and the variability of precipitation predictions was more sensitive to the choice of the cumulus options

rather than micro-physical schemes. Another study conducted by Nasrollahi [42] showed that the best model ability to predict hurricanes was achieved using a particular cumulus parameterization scheme combined with a particular micro-physics scheme. Therefore, the interactions of different physical parameterizations have a considerable impact on model errors, and can be considered as one of the main sources of uncertainty that affect the forecast accuracy.

This paper demonstrates the potential of machine learning techniques to help solve two important problems related to the structural/physical uncertainty in numerical weather prediction models. The first problem addressed herein is the estimation of systematic model errors in output quantities of interest at future times, and the use of this information to improve the model forecasts. The second problem considered is the identification of those specific physical processes that contribute most to the forecast uncertainty in the quantity of interest under specified meteorological conditions.

The application of machine learning techniques to problems in environmental science has grown considerably in recent years. In [18] a kernel based regression method is developed as a forecasting approach with performance close to Ensemble Kalman Filter (EnKF) and less computational resources. Krasnopol et al. [29] employ an Artificial Neural Network technique for developing an ensemble stochastic convection parameterization for climate models. Attia et al. [3] develop a new filtering algorithm called Cluster Hybrid Monte Carlo sampling filter (CLHMC) non-Gaussian data assimilation which relaxes the Gaussian assumptions by employing a clustering step. Moosavi et al. [36] use regression machine learning techniques for adaptive localization in ensemble based data assimilation.

This study is an extension of our work [37] and this focuses on the uncertainty in forecasts of cumulative precipitation caused by imperfect representations of physics and their interaction in the WRF model. The total accumulated precipitation includes all phases of convective and non-convective precipitation. Specifically, we seek to use the discrepancies between WRF forecasts and measured precipitation levels in the past in order to estimate in advance the WRF prediction uncertainty. The model-observation differences contain valuable information about the error dynamics and the missing physics of the model. We use this information to construct two probabilistic functions. The first one maps the discrepancy data and the physical parameters onto the expected forecast errors. The second maps the forecast error levels onto the set of physical parameters that are consistent with them. Both maps are constructed using supervised machine learning techniques, specifically, using Artificial Neural Networks and Random Forests [41]. The two probabilistic maps are used to address the problems posed above, namely the estimation of model errors in output quantities of interest at future times, and the identification of physical processes that contribute most to the forecast uncertainty.

The remainder of this study is organized as follows. Section 2 covers the definition of the model errors. Section 3 describes the proposed approach of error modeling using machine learning. Section 4 reports numerical experiments with the WRF model that illustrate the capability of the new approach to answer two

important questions regarding model errors. Conclusions are drawn in Section 5.

2. Model errors

First-principles computer models capture our knowledge about the physical laws that govern the evolution of a real physical system. The model evolves an initial state at the initial time to states at future times. All models are imperfect, e.g., atmospheric model uncertainties are associated with sub-grid modeling, boundary conditions, and forcings. All these modeling uncertainties are aggregated into a component that is generically called *model error* [19, 45, 46]. In the past decade there has been a considerable scientific effort to incorporate model errors and estimate their impact on the best estimate in both variational and statistical approaches [1, 6, 21, 50, 57, 58, 63].

In what follows, we describe our mathematical formulation of the model error associated with NWP models. A similar formulation has been used in [38] where the model structural uncertainty is studied based on the information provided by the discrepancy between the model solution and the true state of the physical system, as measured by the available observations.

Consider the following NWP computer model $\mathcal{M}$, that describes the time-evolution of the state of the atmosphere:

$$\mathbf{x}_t = \mathcal{M}\left(\mathbf{x}_{t-1}, \Theta\right), \quad t = 1, \cdots, T\,. \tag{1a}$$

The state vector $\mathbf{x}_t \in \mathbb{R}^n$ contains the dynamic variables of the atmosphere such as temperature, pressure, tracer concentrations etc. at all spatial locations covered by the model, and at t. All the physical parameters of the model are lumped into $\Theta \in \mathbb{R}^\ell$.

Formally, the true state of the atmosphere can be described by a physical process $\mathcal{P}$ with internal states υ_t, which are unknown. The atmosphere, as an abstract physical process, evolves in time as follows:

$$\upsilon_t = \mathcal{P}\left(\upsilon_{t-1}\right), \quad t = 1, \cdots, T. \tag{1b}$$

The model state seeks to approximates the physical state:

$$\mathbf{x}_t \approx \psi(\upsilon_t), \quad t = 1, \cdots, T, \tag{1c}$$

where the operator ψ maps the physical space onto the model space, e.g., by sampling the continuous meteorological fields onto a finite dimensional computational grid [38].

Assume that the model state at $t-1$ has the ideal value obtained from the true state via (1c). The model prediction at t will differ from the reality:

$$\psi(\upsilon_t) = \mathcal{M}\big(\psi(\upsilon_{t-1}), \Theta\big) + \boldsymbol{\delta}_t\big(\upsilon_t\big), \quad t = 1, \cdots, T, \tag{2}$$

where the discrepancy $\boldsymbol{\delta}_t \in \mathbb{R}^n$ between the model prediction and reality is the *structural model error*. This vector lives in the model space.

Although the global physical state υ_t is unknown, we obtain information about it by measuring of a finite number of observables $\mathbf{y}_t \in \mathbb{R}^m$, as follows:

$$\mathbf{y}_t = h(\upsilon_t) + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \mathbf{R}_t), \quad t = 1, \cdots, T, \tag{3}$$

Here h is the observation operator that maps the true state of atmosphere to the observation space, and the observation error ϵ_t is assumed to be normally distributed.

In order to relate the model state to observations we also consider the observation operator $\mathcal{H}$ that maps the model state onto the observation space; the model-predicted values $\mathbf{o}_t \in \mathbb{R}^m$ of the observations (3) are:

$$\mathbf{o}_t = \mathcal{H}(\mathbf{x}_t), \quad t = 1, \cdots, T. \tag{4}$$

We note that the measurements $\mathbf{y}_t$ and the predictions $\mathbf{o}_t$ live in the same space and therefore can be directly compared. The difference between the observations of the real system and the model predicted values of these observables (4) represent the model error in observation space:

$$\boldsymbol{\Delta}_t = \mathbf{o}_t - \mathbf{y}_t \in \mathbb{R}^m, \quad t = 1, \cdots, T. \tag{5}$$

For clarity, in what follows we make the following simplifying assumptions [38]:

- the physical system is finite dimensional $\upsilon_t \in \mathbb{R}^n$,
- the model state lives in the same space as reality, i.e., $\mathbf{x}_t \approx \upsilon_t$ and $\psi(\cdot) \equiv id$ is the identity operator in (1c), and
- $\mathcal{H}(\cdot) \equiv h(\cdot)$ in (3) and (4).

These assumptions imply that the discretization errors and representativeness errors are negligbile, and that the main source of error are the parameterized physical processes represented by Θ and the interaction among these processes. Uncertainties from other sources, such as boundary conditions, are also assumed to be small.

With these assumptions, the evolution equations for the physical system (1b) and the physical observations equation (3) become, respectively:

$$\upsilon_t = \mathcal{M}\big(\upsilon_{t-1}, \Theta\big) + \boldsymbol{\delta}_t\big(\upsilon_t\big), \quad t = 1, \cdots, T, \tag{6a}$$

$$\mathbf{y}_t = h(\upsilon_t) + \epsilon_t. \tag{6b}$$

The model errors $\boldsymbol{\delta}_t$ (2) are not fully known at any time t, as having the exact errors is akin to having a perfect model. However, the discrepancies between the modeled and measured observable quantities (5) at past times have been computed and are available at the current time t.

Our goal is to use the errors in observable quantities at past times, $\boldsymbol{\Delta}_\tau$ for $\tau = t-1, t-2, \cdots$, in order to estimate the model error $\boldsymbol{\delta}_\tau$ at future times $\tau = t, t+1, \cdots$. This is achieved by unravelling the hidden information in the past $\boldsymbol{\Delta}_\tau$ values. Good estimates of the discrepancy $\boldsymbol{\delta}_t$, when available, could improve model predictions by applying the correction (6a) to model results:

$$\mathbf{v}_t \approx \mathbf{x}_t + \boldsymbol{\delta}_t. \tag{7}$$

Our proposed error modeling approach constructs input-output mappings to estimate given aspects of model errors $\boldsymbol{\delta}_t$. The inputs to these mappings are the physical parameters Θ of the model. The outputs to these mappings are different aspects of the error in a quantity of interest, such as the model errors over a specific geographical location, or the error norm of model error integrated over the entire domain.

Specifically, the aspect of interest (quantity of interest) in this study is the error in precipitation levels forecasted by the model. The parameters Θ describe the set of physical processes that are essential to be included in the WRF model in order to produce accurate precipitation forecasts. The WRF model is modular and different combinations of the physical packages can be selected, each corresponding to a different value of Θ.

We use the error mappings learned from past model runs to estimate the model errors $\boldsymbol{\delta}_t$. We also consider estimating what combination of physical processes Θ leads to lower model errors, or reversely, what interactions of which physics cause larger errors in the prediction of the quantity of interest.

3. Approximating model errors using machine learning

We propose a multivariate input-output learning model to predict the model errors $\boldsymbol{\delta}$, defined in (2), stemming from the uncertainty in parameters Θ. To this end, we define a probabilistic function ϕ that maps every set of input features $F \in \mathbb{R}^r$ to output target variables $\Lambda \in \mathbb{R}^o$:

$$\phi(F) \approx \Lambda\,, \tag{8}$$

and approximate the function ϕ using machine learning.

Different particular definitions of ϕ in (8) will be used to address two different problems related to model errors, as follows:

1. The first problem is to estimate the systematic model error in certain quantities of interest at future times, and to use this information in order to improve the WRF forecast. To achieve this one quantifies the model error aspects that correspond to running WRF with different physical configurations (different parameters Θ).
2. The second problem is to identify the specific physical processes that contribute most to the forecast uncertainty in the quantity of interest under specified meteorological conditions. To achieve this one finds the model

configurations (physical parameters Θ) that lead to forecast errors smaller that a given threshold under specified meteorological conditions.

In what follows we explain in detail the function ϕ specification, the input features, and the target variables for each of these problems.

3.1. Problem one: estimating in advance aspects of interest of the model error

Forecasts produced by NWP models are contaminated by model errors. These model errors are highly correlated in time; hence historical information about the model errors can be used as an input to the learning model to gain insight about model errors that affect the forecast. We are interested in the uncertainty caused due to the interaction between the various components in the physics based model; these interactions are lumped into the parameter Θ that is supplied as an input to the learning model. The learning model aims to predict the error of NWP model of next forecast window using the historical values of model error and the physical parameters used in the model. We define the following mapping:

$$\phi^{\text{error}}\left(\Theta, \boldsymbol{\Delta}_\tau, \mathbf{o}_\tau, \mathbf{o}_t\right) \approx \boldsymbol{\Delta}_t \quad \tau < t. \tag{9}$$

We use a machine learning algorithm to approximate the function ϕ^{error}. The learning model is trained using a dataset that consists of the following inputs:

- WRF physical packages that affect the physical quantity of interest (Θ),
- historical WRF forecasts ($\mathbf{o}_\tau$ for $\tau \leq t-1$),
- historical model discrepancies ($\boldsymbol{\Delta}_\tau$ for $\tau \leq t-1$),
- WRF forecast at the current time ($\mathbf{o}_t$),
- the available model discrepancy at the current time ($\boldsymbol{\Delta}_t$) since we have access to the observations from reality y_t at the current time step.

In supervised learning process, the learning model identifies the effect of physical packages, the historical WRF forecast, the historical model discrepancy, and the WRF forecast at the current time on the available model discrepancy at the current time. After the model get trained on the historical data, it yields an approximation to the mapping ϕ^{error}. We denote this approximate mapping by $\widehat{\phi}^{\text{error}}$.

During the test phase the approximate mapping $\widehat{\phi}^{\text{error}}$ is used to estimate the model discrepancy $\widehat{\boldsymbol{\Delta}}_{t+1}$ in advance. We emphasize that the model prediction (WRF forecast) at the time of interest $t+1$ ($\mathbf{o}_{t+1}$) is available, where as the model discrepancy $\widehat{\boldsymbol{\Delta}}_{t+1}$ is an unknown quantity. In fact the run time of WRF is much smaller than the time interval between t and $t+1$, or in other way, the time interval is large enough to run the WRF model and obtain the forecast for next time window, estimate the model errors for next time window and finally improve the model forecast by combining the model forecast and model errors.

At the test time we predict the future model error as follows:

$$\widehat{\boldsymbol{\Delta}}_{t+1} \approx \widehat{\phi}^{\text{error}}\left(\Theta, \boldsymbol{\Delta}_{\tau}, \mathbf{o}_{\tau}, \mathbf{o}_{t+1}\right), \quad \tau < t+1\,.$$

As explained in [38], the predicted error $\widehat{\boldsymbol{\Delta}}_{t+1}$ in the observation space can be used to estimate the error $\boldsymbol{\delta}_{t+1}$ in the model space. In order to achieve this one needs to use additional information about the structure of the model and the observation operator. For example, if the error $\widehat{\boldsymbol{\Delta}}_{t+1}$ represents the projection of the full model error onto the observation space, we have:

$$\boldsymbol{\Delta}_{t+1} \approx \mathbf{H}_t \cdot \boldsymbol{\delta}_{t+1}, \quad \widehat{\boldsymbol{\delta}}_{t+1} \approx \mathbf{H}_t \left(\mathbf{H}_t^T \mathbf{H}_t\right)^{-1} \mathbf{H}_t^T \cdot \widehat{\boldsymbol{\Delta}}_{t+1}, \tag{10a}$$

where we use the linearized observation operator at the current time, $\mathbf{H}_t = h'(\boldsymbol{x}_t)$. A more complex approach is to use a Kalman update formula:

$$\widehat{\boldsymbol{\delta}}_{t+1} \approx \text{cov}(\mathbf{x}_t, \mathbf{o}_t) \left(\text{cov}(\mathbf{o}_t, \mathbf{o}_t) + \mathbf{R}_t\right)^{-1} \widehat{\boldsymbol{\Delta}}_{t+1}, \tag{10b}$$

where $\mathbf{R}_t$ is the covariance of observation errors. The Kalman update approach requires estimates of the covariance matrices between model variables; such covariances are already available in an ensemble based data assimilation system. Once we estimate the future model error $\boldsymbol{\delta}_{t+1}$, we can improve the NWP output using equation (7).

3.2. Problem two: identifying the physical packages that contribute most to the forecast uncertainty

Typical NWP models incorporate an array of different physical packages to represent multiple physical phenomena that act simultaneously. Each physical package contains several alternative configurations (e.g., parameterizations or numerical solvers) that affect the accuracy of the forecasts produced by the NWP model. A particular scheme in a certain physical package best captures the reality under some specific conditions (e.g., time of the year, representation of sea-ice, etc.). The primary focus of this study is the accuracy of precipitation forecasts, therefore we seek to learn the impacts of all the physical packages that affect precipitation. To this end, we define the following mapping:

$$\phi^{\text{physics}}\left(\boldsymbol{\Delta}_t\right) \approx \Theta\,, \tag{11}$$

that estimates the configuration Θ of the physical packages such that the WRF run generates a forecast with an error consistent with the prescribed level $\boldsymbol{\Delta}_t$ (where $\boldsymbol{\Delta}_t$ defined in equation (5) is the forecast error in observation space at time t.)

We train the model to learn the effect of the physical schemes on the mismatch between WRF forecasts and reality. The input data required for the training process is obtained by running the model with various physical package configurations Θ_i^{train}, and comparing the model forecast against the observations at all past times τ to obtain the corresponding errors $\boldsymbol{\Delta}_{\tau,i}^{\text{train}}$ for $\tau \leq t$

and $i \in \{training\ data\ set\}$. The output data is the corresponding physical combinations Θ that leads to the input error threshold.

In order to estimate the combinations of physical process configuration that contribute most to the uncertainty in predicting precipitation we take the following approach. The dataset consisting of the observable discrepancies during the current time window $\mathbf{\Delta}_t$ is split into a training part and a testing part. In the test phase we use the machine learned approximation of ϕ^{physics} (denoted by $\widehat{\phi}^{\text{physics}}$) to estimate the physical process settings $\widehat{\Theta}_j^1$ that are consistent with the observable errors $\mathbf{\Delta}_{t,j}^{\{1\}}$. Note that $\widehat{\Theta}_j^1$ is a prediction that is obtained by using $\widehat{\phi}^{\text{physics}}$. Here we select $\mathbf{\Delta}_{t,j}^{\{1\}} = \mathbf{\Delta}_{t,j}^{\text{test}}$ for each $j \in \{test\ data\ set\}$.

Next, we reduce the desired forecast error level to $\mathbf{\Delta}_{t,j}^{\{2\}} = \mathbf{\Delta}_{t,j}^{\{1\}}/2$, and use the approximated function $\widehat{\phi}^{\text{physics}}$ to estimate the physical process setting $\widehat{\Theta}_j^{\{2\}}$ that corresponds to this more accurate forecast. To identify the package setting that has the largest impact on the observable error we monitor the variability in the predicted parameters $\widehat{\Theta}^{\{2\}} - \widehat{\Theta}^{\{1\}}$. Specifically, the number of times the setting of a physical process in $\widehat{\Theta}_j^2$ is different from its setting in $\widehat{\Theta}_j^1$ is an indicator of the variability in model prediction when that package is changed. A higher variability in predicted physical packages implies a larger contribution towards the model errors - as estimated by the ML model.

We want to emphasize that the goal of this approach is not to improve the immediate forecast. The idea is to use the data over long periods and draw conclusions to which packages is the precipitation most sensitive to. These packages may vary over different seasons and regions. Having this data, will guide us to improve the "right" physical package according to the regions and the time of the year. For best practical use, this approach has to be used in conjunction with large datasets across seasons and regions to determine the packages that need to be improved.

3.3. Machine learning algorithms

In order to approximate the functions (9) and (11) discussed earlier we use regression machine learning methods. Choosing a right learning algorithm to use is challenging as it largely depends on the problem and the data available [3, 2, 39, 36]. Here, we use Random Forests (RF) and Artificial Neural Networks (ANN) as our learning algorithms [41]. Both RF and ANN algorithms tan handle non-linearity in regression and classification. Given that the physical phenomena governing precipitation are highly nonlinear, and and atmospheric dynamics is chaotic, we believe that RF and ANN approaches are well suited to capture the associated features. We briefly review these techniques next.

3.3.1. Random forests

A random forest [5] is an ensemble based method that constructs multiple decision trees. The principle idea behind ensemble methods is that a group of weak learners can come together to form a strong learner [4, 5]. The decision tree is built top-down from observations of target variables. The observation

dataset is partitioned, smaller subsets are represented in branches, and decisions about the target variables are represented in the leaves.

There are many specific decision-tree algorithms available, including ID3 (Iterative Dichotomiser 3) [48], C4.5 (successor of ID3) [49], CART (Classification And Regression Tree), CHAID (CHi-squared Automatic Interaction Detector), and conditional inference trees [54]. If the dataset has multiple attributes, one can decide which attribute to place at the root or at different levels of the tree by considering different criteria such as information gain or the gini index [7].

Trees can be non-robust, with small changes in the tree leading to large changes in regression results. Moreover, trees tend to over-fit the data [52]. The random forest algorithm uses the bagging technique for building an ensemble of decision trees which are accurate and powerful at handling large, high dimensional datasets. Moreover, the bagging technique greatly reduces the variance [10]. For each tree in the forest, a bootstrap sample [4, 10] is selected from the dataset and instead of examining all possible feature-splits, some subset of the features is selected [32]. The node then splits on the best feature in the subset. By using a random sample of features the correlation between trees in the ensemble decreases, and the learning for each tree is much faster by restricting the features considered for each node.

Gradient boosting is yet another popular algorithm based on decision trees. The gradient boosting algorithm differs from random forests in the way it builds the ensemble of trees and combines the results from them. Instead of building the decision trees independently, as in the case of random forests, gradient boosting algorithm builds one tree at a time in an additive, forward stage-wise manner. Also, instead of combining the results at the end of the process, the gradient boosting combines the results along the way [17]. Gradient boosting may not be a good choice when there is a lot of noise in the data and additionally they tend to be harder to tune than random forests.

3.3.2. Artificial neural networks

ANN is a computational model inspired by human brain's biological structure. ANN consist of neurons and connections between the neurons (weights) which are organized in layers. At least three layers of neurons (an input layer, a hidden layer, and an output layer.) are required for construction of most neural networks (A single layered perceptron with no hidden layer is a notable exception). The input layer distributes the input signals to the first hidden layer. The feed-forward operation in a network passes information to neurons in a subsequent hidden layer. The neurons combine this information, and the output of each layer is obtained by passing the combined information through a differentiable transfer function that can be log-sigmoid, hyperbolic tangent sigmoid, or linear transfer function.

In supervised learning the network is provided with samples from which it discovers the relations of inputs and outputs. The learning problem consists of finding the optimal parameters of network such that the error between the desired output and the output signal of the network is minimized. The network

first is initialized with randomly chosen weights and then the error is back-propagated through the network using a gradient descent method. The gradient of the error function is computed and used to modify weights and biases such that the error between the desired output and the output signal of the network is minimized [15, 51] . This process is repeated iteratively until the network output is close to the desired output [22].

4. Numerical experiments

We apply the proposed learning models to the Weather Research and Forecasting model [61] in order to:

- predict the bias in precipitation forecast caused by structural model errors,
- predict the statistics associated with the precipitation errors, and
- identify the specific physics packages that contribute most to precipitation forecast errors for given meteorological conditions.

4.1. The WRF model

In this study we use the non-hydrostatic WRF model version 3.3. The simulation domain covers the continental United States and has dimensions of 60×73 horizontal grid points in the west-east and south-north directions respectively, with a horizontal grid spacing of $60km$ [59]. The grid has 60 vertical levels to cover the troposphere and lower part of the stratosphere between the surface to approximately $20km$. In all simulations, the six-hourly analysis from the National Centers for Environmental Prediction (NCEP) are used as the initial and boundary conditions of the model [44]. By NCEP six-hourly analysis we mean the analysis provided by North American Mesoscale model. We use WRF processing system to convert this six-hourly NCEP analysis to process initial and lateral boundary conditions. The NCEP stage IV analysis estimates are available at an hourly temporal resolution over continental United States. NCEP stage IV product is a near-real-time product that is generated at NCEP separately based on the NEXRAD Precipitation Processing System [14] and the NWS River Forecast Center (RFC) precipitation processing [53]. For experimental purposes, we use the stage IV NCEP analysis as a proxy for the true state of the atmosphere and the WRF forecasts will be evaluated against the stage IV NCEP analysis. The simulation window begins at 6AM UTC (Universal Time Coordinated) on May 1st 2017, and the simulation time is a six hour window time the same day. The "true" states of the atmosphere are available in the form of hourly stage IV NCEP data. All the numerical experiments use the NCEP analysis data to run WRF model on May 1st 2017.

The model configuration parameters Θ represent various combinations of micro-physics schemes, cumulus parameterizations, short wave, and long wave radiation schemes. We use the numerical values detailed in WRF model physics options and references [62] to represent the physics parametrizations. The

micro-physics option provides atmospheric heat and moisture tendencies in atmosphere which also accounts for the vertical flux of precipitation and the sedimentation process. The cumulus parameterization is used to vertically redistribute heat and moisture independent of latent heating due to precipitation. The long wave radiation considers clear-sky and cloud upward and downward radiation fluxes and the short wave radiation considers clear-sky and cloudy solar fluxes.

A total number of 252 combinations of the four physical modules are used in the simulations. The micro-physics schemes include: Kessler [28], Lin [33], WSM3 Hong [23], WSM5 Hong [23], Eta (Ferrier), WSM6 [24], Goddard [55], Thompson [56], Morrison [40]. The cumulus physics schemes applied are: Kain-Fritsch [27], Betts-Miller-Janjic [25], Grell Freitas[20]. The long wave radiation physics include: RRTM [35], CAM [9]. Short wave radiation physics include: Dudhia [11], Goddard [8], CAM [9].

For each of the 252 different physics combinations, the effect of each physics combination on precipitation is investigated. The stage IV NCEP analysis grid points are 428×614, while the WRF computational model have 60×73 grid points. For obtaining the discrepancy between the WRF forecast and stage IV NCEP analysis we linearly interpolate the analysis to transfer the physical variables onto the model grid. To project the stage IV onto WRF computational grid we use the following interpolation scheme:

- For every grid point on the computational grid, find six nearest neighbours using great-circle distance as the metric.
- Use the weighted average to approximate the accumulated precipitation at the grid point of interest.

Given that stage IV NCEP grid is substantially finer than the WRF computational grid used by us, the above scheme is a reasonable means to project from stage IV grid to WRF computational grid. Figure 1 and 2 shows the WRF forecast and NCEP stage IV analysis at 12PM, 5/1/2017. The NCEP stage IV analysis is used as a proxy for truth and the WRF forecast is compared against the NCEP stage IV analysis to assess the errors in forecast. The forecast in Figure 1 corresponds to the following physics settings: micro-physics: Kessler, cu-physics: Kain-Fritsch, ra-lw-physics: CAM , ra-sw-physics: Dudhia. Figures 3 and 4 show contours of discrepancy in accumulated precipitation between the truth and forecast. The cumulative precipitation is calculated from 6AM to 12PM ($\mathbf{\Delta}_{t=12\text{PM}}$) discussed in equation (5) for two different physical combinations, which illustrates the effect that changing the physical schemes has on the forecast.

We demonstrate our learning algorithms to forecast precipitation in the state of Virginia on May 1st 2017 at 6PM. Our goal is to use the learning algorithms to correct the bias created due to model errors and hence improve the forecast for precipitation. As described in section 3.1, we learn the function ϕ^{error} of equation (9) using the training data from the previous forecast window (6AM to

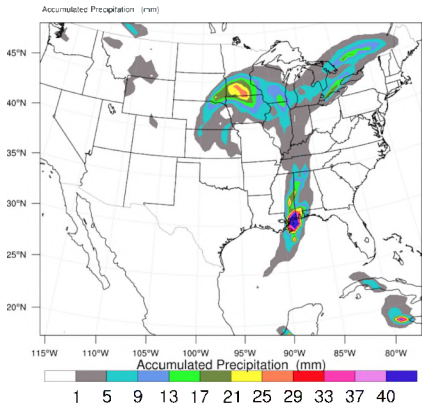


Figure 1: NCEP stage IV analysis at 12PM provides a proxy for the true state of the atmosphere

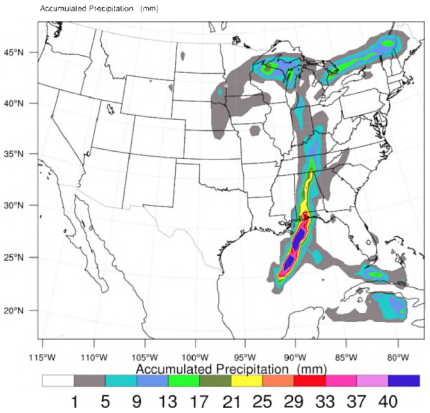


Figure 2: WRF forecast at 12PM corresponding to the physics micro-physics: Kessler, cumulus physics: Kain-Fritsch, long wave radiation physics: Cam, short wave radiation physics: Dudhia

12PM):

$$\phi^{\text{error}}\left(\Theta, \boldsymbol{\Delta}_{\tau}, \mathbf{o}_{\tau}, \mathbf{o}_{t-12\text{PM}}\right) \approx \boldsymbol{\Delta}_{t-12\text{PM}}, \quad 6\text{AM} \leq \tau < 12\text{PM}.$$

We use two learning algorithms to approximate the function ϕ^{error}, namely, the RF with ten trees and CART learning tree algorithm in the forest and an ANN with four hidden layers and hyperbolic tangent sigmoid activation function in each layer are employed using Scikit-learn, machine learning library in Python [47]. For training purposes, we use the NCEP analysis of the May 1st 2017 at 6AM as initial conditions for the WRF model. The forecast window is 6 hours and the WRF model forecast final simulation time is 12PM. The input features are:

- The physics combinations (Θ).
- The hourly WRF forecasts projected onto observation space o_{τ}, $6\text{AM} \leq \tau \leq 12\text{PM}$. The WRF state ($\mathbf{x}_t$) includes all model variables such as temperature, zonal winds (meridional, zonal, and vertical velocities), geopotential, and surface pressure. The observation operator computes the precipitation portion of the WRF output vector, $\mathbf{o}_t \equiv \mathbf{x}_t^{\text{precipitation}}$. Accordingly, $\boldsymbol{\Delta}_t$ is the discrepancy between WRF precipitation forecast $\mathbf{o}_t$ and the observed precipitation $\mathbf{y}_t$.
- The observed discrepancies at past times ($\boldsymbol{\Delta}_{\tau}$, $6\text{AM} \leq \tau < 12\text{PM}$).

The output variable is the discrepancy between the NCEP stage IV analysis and the WRF forecast at 12PM, i.e., the observable discrepancies for the current forecast window ($\boldsymbol{\Delta}_{t=12\text{PM}}$). In fact, for each of the 252 different physical

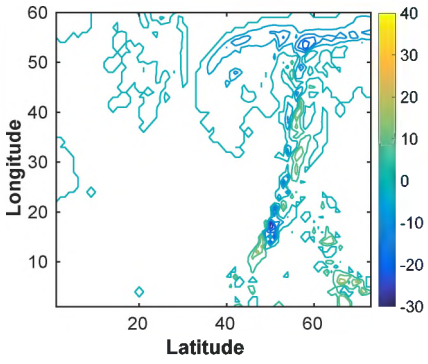


Figure 3: Discrepancy between the forecasted accumulated precipitation from 6 AM to 12 PM and the "truth" (that is NCEP stage IV analysis). The parameterizations used are microphysics scheme: Kessler, cumulus physics: Kain-Fritsch, short wave radiation: CAM, long wave radiation: Dudhia

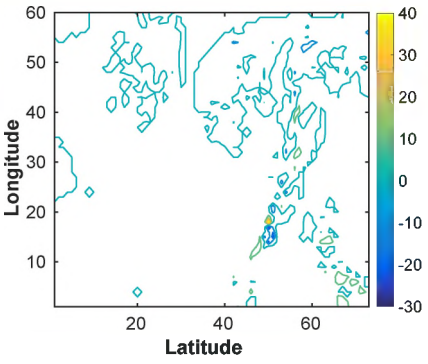


Figure 4: Discrepancy between the forecasted accumulated precipitation from 6 AM to 12 PM and the "truth" (that is NCEP stage IV analysis). The parameterizations used are microphysics scheme: Lin, cumulus physics: Kain-Fritsch, short wave radiation: RRTM Mlawer, long wave radiation: CAM

configurations, the WRF model forecast as well as the difference between the WRF forecast and the analysis are provided as input-output combinations for learning the function ϕ^{error}. The number of grid points over the state of Virginia is 14×12. Therefore for each physical combination we have 168 grid points, and the total number of samples in the training data set is $252 \times 168 = 42,336$ with 15 features. Some example features are physics combinations, WRF output, discrepancy between precipitation forecast and observed precipitation etc.

Both ANN and RF are trained with the above input-output combinations described above and during the training phase, the learning model learns the effect of interaction between different physical configurations on the WRF forecast and model error and obtains the approximation to the function ϕ^{error} which we denote by $\widehat{\phi}^{\text{error}}$. The goal is to have more accurate forecast in the future time windows. We don't have the analysis data of future time windows but we can run WRF for future time windows and also predict the future model error using the approximated function $\widehat{\phi}^{\text{error}}$. Once we obtain the predicted model error we can use that information in order to raise the accuracy of WRF forecast. In the testing phase we use the function $\widehat{\phi}^{\text{error}}$ to predict the future forecast error $\widehat{\mathbf{\Delta}}_{t=6\text{PM}}$ given the combination of physical parameters as well as the WRF forecast at time 6PM as input features.

$$\widehat{\mathbf{\Delta}}_{t=6\text{PM}} \approx \widehat{\phi}^{\text{error}}\left(\Theta, \mathbf{\Delta}_{\tau}, \mathbf{o}_{\tau}, \mathbf{o}_{t=6\text{PM}}\right), \quad 12\text{PM} \leq \tau < 6\text{PM}.$$

To quantify the accuracy of the predicted error we calculate the Root Mean

Squared Error (RMSE) between the true and predicted discrepancies at 6PM:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{\boldsymbol{\Delta}}^{i}_{t=6\text{PM}} - \boldsymbol{\Delta}^{i}_{t=6\text{PM}}\right)^2}, \tag{12}$$

where $n = 168$ is the number of grid points over Virginia,. $\widehat{\boldsymbol{\Delta}}^{i}_{t=6\text{PM}}$ is the predicted discrepancy in the i^{th} grid point, and $\boldsymbol{\Delta}^{i}_{t=6\text{PM}}$ is the i^{th} actual discrepancy in the i^{th} grid point. The actual discrepancy is obtained as the difference between the NCEP stage IV analysis and WRF forecast at time $t = 6\text{PM}$. This error metric is computed for each of the 252 different configurations of the physics. The minimum, maximum and average RMSE over the 252 runs is reported in Table 1.

	minimum($RMSE$)	average($RMSE$)	maximum($RMSE$)
ANN	1.264×10^{-3}	1.343×10^{-3}	5.212×10^{-3}
RF	1.841×10^{-3}	1.931×10^{-3}	7.9×10^{-3}

Table 1: The minimum, average, and maximum RMSE between the predicted $\widehat{\boldsymbol{\Delta}}_{t=6\text{PM}}$ and the true $\boldsymbol{\Delta}_{t=6}$ over 252 physics combinations.

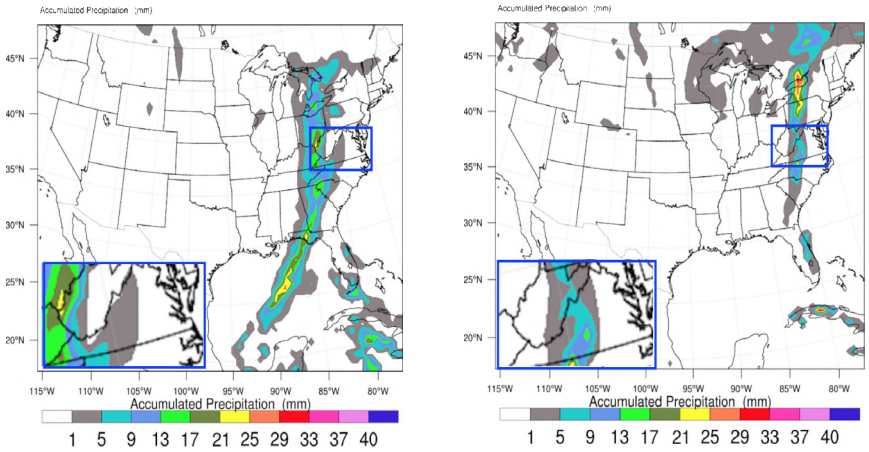
The predicted discrepancy in the observation space $\widehat{\boldsymbol{\Delta}}_{t=6\text{PM}}$ can be used to approximate the discrepancy in the model space $\widehat{\boldsymbol{\delta}}_{t=6\text{PM}}$ using equation (10). Here all the grid points are observed and therefore the error in the model space equal to the error in the observation space. Next, the estimate forecast error can be used to correct the forecast bias caused by model errors using (7), and hence to improve the forecast at 6PM: $\widehat{\mathbf{x}}_{t=6\text{PM}} = \mathbf{x}_{t=6\text{PM}} + \widehat{\boldsymbol{\delta}}_{t=6\text{PM}}$. Figure 5(a) shows the WRF forecast for 6PM for the state of Virginia using the following physics packages [42] (the physics options are given in parentheses):

- Micro-physics (Kessler),
- Cumulus-physics (Kain),
- Short-wave radiation physics (Dudhia),
- Long-wave radiation physics (RRTM).

Figure 5(b) shows the NCEP stage IV analysis at time 6PM, which is our proxy for the true state of the atmosphere. The discrepancy between the NCEP stage IV analysis and the raw WRF forecast is shown in the Figure 6(a). Using the model error prediction we can improve the WRF result by adding the predicted bias to the WRF forecast. The discrepancy between the corrected WRF forecast and the NCEP stage IV analysis is shown in the Figure 6(b). The results show a considerable reduction of model errors as compared to the uncorrected forecast of Figure 6(a). Table 2 shows the minimum and average of original model error vs the improved model errors.

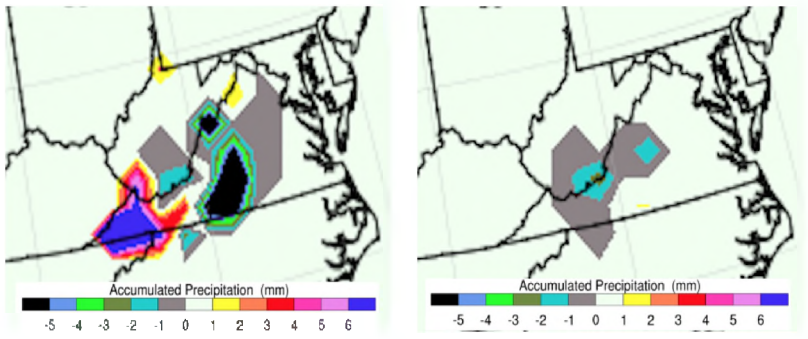
	minimum($\mathbf{\Delta}_{t=6\text{PM}}$)	average($\mathbf{\Delta}_{t=6\text{PM}}$)
Original forecast	6.751×10^{-2}	5.025×10^{-1}
Improved forecast	2.134×10^{-4}	6.352×10^{-2}

Table 2: The minimum and average of $\mathbf{\Delta}_{t\ \ 6\text{PM}}$ for the original WRF forecast vs the improved forecast



(a) Original WRF prediction (b) NCEP Stage IV analysis

Figure 5: WRF prediction and NCEP Stage IV analysis at 6PM on 5/1/2017. Zoom-in panels show the predictions over Virginia.



(a) Discrepancy between original WRF forecast and NCEP stage IV analysis

(b) Discrepancy between the corrected WRF forecast and the NCEP stage IV analysis

Figure 6: Discrepancy between WRF forecasts and the NCEP stage IV analysis over Virginia at 6PM on 5/1/2017. The forecast correction clearly improves the model results.

4.2. Experiments for problem one: predicting the norm of precipitation forecast error over the entire domain

We now seek to estimate the two-norm of precipitation model error over the entire continental U.S., which gives a global metric for the accuracy of the WRF forecast, and helps provide insight about the physics configurations that result in more accurate forecasts. To this end the following mapping is constructed:

$$\phi^{\text{error}}\left(\Theta, \|\mathbf{o}_\tau\|_2, \|\boldsymbol{\Delta}_\tau\|_2, \|\mathbf{o}_{t=12\text{PM}}\|_2, \bar{\mathbf{o}}_{t=12\text{PM}}\right) \approx \|\boldsymbol{\Delta}_{t=12\text{PM}}\|_2, \quad 6\text{AM} \leq \tau < 12\text{PM}.$$

To build the training dataset, we run WRF with each of the 252 different physical configurations. The forecast window is 6 hours and the WRF model forecast final simulation time is at 12PM. The hourly WRF forecast and discrepancy between the stage IV analysis and WRF forecast is used as training features.

The input features are:

- different physics schemes (Θ),
- the norms of the WRF model predictions at previous time windows, as well as at the current time ($\|\mathbf{o}_{t=12\text{PM}}\|_2, \|\mathbf{o}_\tau\|_2$, $6\text{AM} \leq \tau < 12\text{PM}$), and
- the norms of past observed discrepancies ($\|\boldsymbol{\Delta}_\tau\|_2$, $6\text{AM} \leq \tau < 12\text{PM}$).

The output variable is the norm of the discrepancy between WRF precipitation prediction and the accumulated precipitation in NCEP stage IV analysis for the current time window ($\|\boldsymbol{\Delta}_{t=12\text{PM}}\|_2$).

We use two different learning algorithms, namely, RF with ten trees in the forest and ANN with four hidden layers and hyperbolic tangent sigmoid activation function in each layer. The total number of samples in the training set is 252 with 15 of features. During the training phase the model learns the effect of interaction of different physical configurations on model error and obtains the approximated function $\widehat{\phi}^{\text{error}}$.

In the test phase we feed the approximated function the model information from 1PM to the endpoint of the next forecast window 6PM to predict the norm of the model error $\|\widehat{\boldsymbol{\Delta}}_{t=6\text{PM}}\|_2$.

$$\widehat{\phi}^{\text{error}}\left(\Theta, \|\mathbf{o}_\tau\|_2, \|\boldsymbol{\Delta}_\tau\|_2, \|\mathbf{o}_{t=6\text{PM}}\|_2, \bar{\mathbf{o}}_{t=6\text{PM}}\right) \approx \|\boldsymbol{\Delta}_{t=6\text{PM}}\|_2, \quad 12\text{PM} \leq \tau < 6\text{PM}.$$

Validation of the learned error mapping. Table 3 shows the RMSE between the actual and predicted norms of discrepancies for ANN and RF. The RMSE is taken over the 252 runs with different physics combinations. Both learning models perform well, with the ANN giving slightly better results than the RF.

	$RMSE(\|\widehat{\boldsymbol{\Delta}}_{t=6\text{PM}}\|_2, \|\boldsymbol{\Delta}_{t=6\text{PM}}\|_2)$
ANN	2.6109×10^{-3}
RF	2.9188×10^{-3}

Table 3: Difference between predicted discrepancy norm $\|\widehat{\boldsymbol{\Delta}}_{t=6\text{PM}}\|_2$ and the reference discrepancy norm $\|\boldsymbol{\Delta}_{t=6\text{PM}}\|_2$. The $RMSE$ is taken over all test cases.

Analysis of the best combination of physical packages. Based on our prediction of the norm of model error, the best physics combination that leads to lowest norm of precipitation error over the entire continental U.S. for the given meteorological conditions is:

- the BMJ cumulus parameterization, combined with
- the WSM5 micro-physics,
- Cam long wave, and
- Dudhia short wave radiation physics.

According to the true model errors, the best physics combination leading to the lowest norm of model error is achieved using the BMJ cumulus parameterization, combined with the WSM5 micro-physics, Cam long wave, and Cam short wave radiation physics. We have tabulated the norms of the model errors in 4

	$\|\boldsymbol{\Delta}_{t=6\text{PM}}\|_2$
Best physics combination	1.512×10^{-2}
Predicted physics combination	1.625×10^{-2}

Table 4: Norms of the model error for best physics combination and predicted physics combination.

4.3. Experiments for problem two: identify the physical processes that contribute most to the forecast uncertainty

The interaction of different physical processes greatly affects precipitation forecast, and we are interested in identifying the major sources of model errors in WRF. To this end we construct the physics mapping (11) using the norm and the statistical characteristics of the model-data discrepancy (over the entire U.S.) as input features:

$$\phi^{\text{physics}}\left(\bar{\boldsymbol{\Delta}}_{t=12\text{PM}}, \|\boldsymbol{\Delta}_{t=12\text{PM}}\|_2\right) \approx \Theta.$$

Statistical characteristics include the mean, minimum, maximum, and variance of the filed across all grid points over the continental U.S. Note that this is slightly different than (11) where the inputs are the raw values of these discrepancies for each grid point. The output variable is the combination of physical

processes Θ that leads to model errors consistent with the input pattern $\bar{\mathbf{\Delta}}_{t=12\text{PM}}$ and $\|\mathbf{\Delta}_{t=12\text{PM}}\|_2$.

To build the dataset, the WRF model is simulated for each of the 252 different physical configurations, and the mismatches between the WRF forecasts and the NCEP stage IV analysis at the end of the current forecast window are obtained. Similar to the previous experiment, the initial conditions used in the WRF model is the NCEP analysis for the May 1st 2017 at 6AM. The forecast window is 6 hours and the WRF model forecast is obtained for time 12PM. The discrepancy between the NCEP stage IV analysis at 12PM and WRF forecast at 12PM forms the observable discrepancy for the current forecast window $\mathbf{\Delta}_{t=12\text{PM}}$. For each of the 252 different physical configurations, this process is repeated and statistical characteristics of the WRF forecast model error $\bar{\mathbf{\Delta}}_{t=12\text{PM}}$, and the norm of model error $\|\mathbf{\Delta}_{t=12\text{PM}}\|_2$ are used as feature values of the function ϕ^{physics}.

Validation of the learned physics mapping. From all the collected data points, 80% (202 samples) are used for training the learning model, and the remaining 20% (50 samples) are used for testing purposes. The learning model uses the training dataset to learn the approximate mapping $\widehat{\phi}^{\text{physics}}$. This function is applied to the each of the 50 test samples $\mathbf{\Delta}^{\text{test}}_{t=12\text{PM}}$ to obtain the predicted physical combinations $\widehat{\Theta}_1$. In order to evaluate these predictions, we run the WRF model again with the $\widehat{\Theta}_1$ physical setting and obtain the new forecast $\widehat{\mathbf{o}}_{t=12\text{PM}}$, and the corresponding observable discrepancy $\widehat{\mathbf{\Delta}}^{\text{test}}_{t=12\text{PM}}$. The RMSE between the norm of actual observable discrepancies and the norm of predicted discrepancies are shown in Table 5. The small values of the difference demonstrates the performance of the learning algorithm.

	$RMSE(\|\widehat{\mathbf{\Delta}}^{\text{test}}_{t=12\text{PM}}\|_2, \|\mathbf{\Delta}^{\text{test}}_{t=12\text{PM}}\|_2)$
ANN	4.1376×10^{-3}
RF	5.8214×10^{-3}

Table 5: The RMSE between estimated discrepancy using predicted physical combinations $\widehat{\mathbf{\Delta}}^{\text{test}}_{t=12\text{PM}}$ and the reference discrepancy $\mathbf{\Delta}^{\text{test}}_{t=12\text{PM}}$.

Analysis of variability in physical settings. We repeat the test phase for each of the 50 test samples with the scaled values of observable discrepancies $\mathbf{\Delta}^{\text{test}}_{t=12\text{PM}}/2$ as inputs, and obtain the predicted physical combinations $\widehat{\Theta}_2$. Large variability in the predicted physical settings $\widehat{\Theta}$ indicate that the respective physical packages variability have a strong influence on the WRF forecast error. We count the number of times the predicted physics $\widehat{\Theta}_2$ is different from $\widehat{\Theta}_1$ when the input data spans the entire test data set.

The results shown in Figure 7 indicate that micro-physics and cumulus physics are not too sensitive to the change of input data, whereas short-wave and long-wave radiation physics are quite sensitive to changes in the input data. Therefore our learning model indicates that having an accurate short-wave and

long-wave radiation physics package will aid in greatly reducing the uncertainty in precipitation forecasts due to missing/incorrect physics. These results reinforce the conclusions by Li and Navon in [31], where they introduced an adjoint technique to study sensitivity of the Earth's radiation budget (ERB) to cloud cover, water vapor, atmospheric temperature, and the Earth's surface temperature. They used this adjoint-based technique to calculate the sensitivities the ERB to cloud cover, water vapor etc. They also found that the outgoing long-wave radiation (OLR) was one order of magnitude more sensitive to water vapor mixing ratios in the upper troposphere than to those in the middle and lower troposphere. Both uncertainty quantification and sensitivity analysis indicate a strong relation between water variables and radiation physics.

4.4. Some comments on CPU costs for the methods

Bulk of the computational time is spent in running the WRF models and training the data. For running the WRF models we used a workstation with 40 threads (two cores, ten processor per core with hyperthreading) and 128 GB of RAM. A single model run required about five minutes on the aforementioned workstation. The dataset was trained on a laptop. The training times for RF was higher than that for ANN. For the bias correction problem, it took about three hours to train the RF and about two hours to train the ANN. For the prediction of statistics associated with the precipitation errors, both RF and ANN algorithms required less than an hour of training. Note that all the ML algorithms were run on a laptop and these timings can be improved dramatically with optimized code running on larger machines. The testing phase consumed negligible time for all experiments when compared to times consumed by training and WRF model runs.

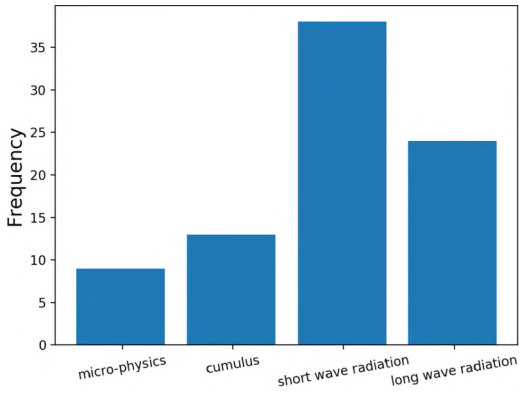


Figure 7: Frequency of change in the physics with respect to change in the input data from $\boldsymbol{\Delta}^{\text{test}}_{t-12\text{PM}}$ to $\boldsymbol{\Delta}^{\text{test}}_{t-12\text{PM}}/2$. Each data set contains 50 data points, and we report here the number of changes of each package.

5. Conclusions

This study proposes a novel use of machine learning techniques to understand, predict, and reduce the uncertainty in the WRF model precipitation forecasts due to the interaction of several physical processes included in the model.

We construct probabilistic approaches to learn the relationships between the configuration of the physical processes used in the simulation and the observed model forecast errors. These relationships are then used to solve two important problems related to model errors, as follows: estimating the systematic model error in a quantity of interest at future times, and identifying the physical processes that contribute most to the forecast uncertainty in a given quantity of interest under specified conditions.

Numerical experiments are carried out with the WRF model using the NCEP analyses as a proxy for the real state of the atmosphere. Ensembles of model runs with different parameter configurations are used to generate the training data. Random forests and Artificial neural network models are used to learn the relationships between physical processes and forecast errors. The experiments validate the new approach, and illustrates how it is able to estimate model errors, indicate best model configurations, and pinpoint to those physical packages that influence most the WRF prediction accuracy.

While the numerical experiments are done with WRF, and are focused on forecasting precipitation, the methodology developed herein is general and can be applied to the study of errors in other models, for other quantities of interest, and for learning additional relationships between model physics and model errors.

In this study we used the model error data from 6 AM to 12PM to predict and understand the model errors from 12 PM to 6PM. This experimental design may suffer from diurnal effects. It is perhaps better to use data from both previous window and previous day to mitigate diurnal effects; we intend to explore this aspect in our future studies. Additionally, as part of our future work, we will explore other advanced machine learning algorithms that fall under the broad category of recurrent neural nets (such as LSTM, GRU, and CNN) and are known to capture the spatial and temporal correlations well to reduce the uncertainty in medium and long-term forecasts.

Acknowledgments

This work was supported in part by the projects AFOSR DDDAS 15RT1037 and AFOSR Computational Mathematics FA9550-17-1-0205 and by the Computational Science Laboratory at Virginia Tech. This material also was based upon work supported by the U.S. Department of Energy, Office of Science, under contract DE-AC02-06CH11357. The authors would like to thank Dr. Răzvan Ştefănescu for his valuable assistance and suggestions regarding WRF runs and the NCEP dataset. The authors would like to thank anonymous reviewers for providing valuable feedback that has helped improve the paper.

References

[1] Santha Akella and Ionel M Navon. Different approaches to model error formulation in 4D-Var: a study with high-resolution advection schemes. *Tellus A*, 61(1):112–128, 2009.

[2] Elham Asgari and Kaveh Bastani. The utility of Hierarchical Dirichlet Processfor relationship detection of latent constructs. In *Academy of Management Proceedings*, 2017.

[3] Ahmed Attia, Azam Moosavi, and Adrian Sandu. Cluster sampling filters for non-Gaussian data assimilation. *arXiv preprint arXiv:1607.03592*, 2016.

[4] Leo Breiman. Bagging predictors. *Machine learning*, 24(2):123–140, 1996.

[5] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[6] Carla Cardinali, Nedjeljka Žagar, Gabor Radnoti, and Roberto Buizza. Representing model error in ensemble data assimilation. *Nonlinear Processes in Geophysics*, 21(5):971–985, 2014.

[7] Lidia Ceriani and Paolo Verme. The origins of the Gini index: extracts from variabilità e mutabilità (1912) by Corrado Gini. *The Journal of Economic Inequality*, 10(3):421–443, 2012.

[8] Ming-Dah Chou and Max J Suarez. A solar radiation parameterization (clirad-sw) for atmospheric studies. *NASA Tech. Memo*, 10460:48, 1999.

[9] Andrew J Conley, Rolando Garcia, Doug Kinnison, Jean-Francois Lamarque, Dan Marsh, Mike Mills, Anne K Smith, Simone Tilmes, Francis Vitt, Hugh Morrison, et al. Description of the NCAR community atmosphere model (CAM 5.0). *NCAR technical note*, 2012.

[10] Thomas G Dietterich et al. Ensemble methods in machine learning. *Multiple classifier systems*, 1857:1–15, 2000.

[11] Jimy Dudhia. Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. *Journal of the Atmospheric Sciences*, 46(20):3077–3107, 1989.

[12] Robert G Fovell. Impact of microphysics on hurricane track and intensity forecasts. In *Preprints, 7th WRF Users' Workshop, NCAR*, 2006.

[13] Robert G Fovell. Influence of cloud-radiative feedback on tropical cyclone motion. In *29th Conference on Hurricanes and Tropical Meteorology*, 2010.

[14] Richard A Fulton, Jay P Breidenbach, Dong-Jun Seo, Dennis A Miller, and Timothy O'Bannon. The wsr-88d rainfall algorithm. *Weather and forecasting*, 13(2):377–395, 1998.

[15] Ken-Ichi Funahashi. On the approximate realization of continuous mappings by neural networks. *Neural networks*, 2(3):183–192, 1989.

[16] WA Gallus Jr. Eta simulations of three extreme rainfall events: Impact of resolution and choice of convective scheme. *Wea. Forecasting*, 14:405–426, 1999.

[17] Aurélien Géron. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems.* O'Reilly Media, 2019.

[18] Robin C Gilbert, Michael B Richman, Theodore B Trafalis, and Lance M Leslie. Machine learning methods for data assimilation. *Computational Intelligence in Architecturing Complex Engineering Systems*, pages 105–112, 2010.

[19] J. Glimm, S. Hou, Y.H. Lee, D.H. Sharp, and K. Ye. Sources of uncertainty and error in the simulation of flow in porous media. *Computational & Applied Mathematics*, 23:109–120, 2004.

[20] Georg A Grell and Saulo R Freitas. A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling. *Atmospheric Chemistry & Physics Discussions*, 13(9), 2013.

[21] James A Hansen. Accounting for model error in ensemble-based state estimation and forecasting. *Monthly Weather Review*, 130(10):2373–2391, 2002.

[22] S.S. Haykin. *Neural Networks and Learning Machines.* Number v. 10 in Neural networks and learning machines. Prentice Hall, 2009.

[23] Song-You Hong, Jimy Dudhia, and Shu-Hua Chen. A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation. *Monthly Weather Review*, 132(1):103–120, 2004.

[24] Song-You Hong and Jeong-Ock Jade Lim. The WRF single-moment 6-class microphysics scheme (wsm6). *J. Korean Meteor. Soc*, 42(2):129–151, 2006.

[25] Zaviša I Janjić. The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Monthly Weather Review*, 122(5):927–945, 1994.

[26] Isidora Jankov, Paul J Schultz, Christopher J Anderson, and Steven E Koch. The impact of different physical parameterizations and their interactions on cold season QPF in the American River basin. *Journal of Hydrometeorology*, 8(5):1141–1151, 2007.

[27] John S Kain. The Kain–Fritsch convective parameterization: an update. *Journal of Applied Meteorology*, 43(1):170–181, 2004.

[28] E Kessler. On the continuity and distribution of water substance in atmospheric circulations. *Atmospheric research*, 38(1-4):109–145, 1995.

[29] VM Krasnopol sky, Michael Fox-Rabinovitz, Alexei Belochitski, Philip J Rasch, Peter Blossey, and Yefim Kogan. *Development of neural network convection parameterizations for climate and NWP models using Cloud Resolving Model simulations*. US Department of Commerce, National Oceanic and Atmospheric Administration, National Weather Service, National Centers for Environmental Prediction, 2011.

[30] François-Xavier Le Dimet and Olivier Talagrand. Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus A: Dynamic Meteorology and Oceanography*, 38(2):97–110, 1986.

[31] Zhijin Li and IM Navon. Sensitivity analysis of outgoing radiation at the top of the atmosphere in the ncep/mrf model. *Journal of Geophysical Research-Atmospheres*, 103(D4):3801–3814, 1998.

[32] Andy Liaw, Matthew Wiener, et al. Classification and regression by random forest. *R news*, 2(3):18–22, 2002.

[33] Yuh-Lang Lin, Richard D Farley, and Harold D Orville. Bulk parameterization of the snow field in a cloud model. *Journal of Climate and Applied Meteorology*, 22(6):1065–1092, 1983.

[34] Marla R Knebl Lowrey and Zong-Liang Yang. Assessing the capability of a regional-scale weather model to simulate extreme precipitation patterns and flooding in central Texas. *Weather and Forecasting*, 23(6):1102–1126, 2008.

[35] Eli J Mlawer, Steven J Taubman, Patrick D Brown, Michael J Iacono, and Shepard A Clough. Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *Journal of Geophysical Research: Atmospheres*, 102(D14):16663–16682, 1997.

[36] Azam Moosavi, Ahmed Attia, and Adrian Sandu. A machine learning approach to adaptive covariance localization. *arXiv preprint arXiv:1801.00548*, 2018.

[37] Azam Moosavi, Vishwas Rao, and Adrian Sandu. A learning-based approach for uncertainty analysis in numerical weather prediction models. In João M. F. Rodrigues, Pedro J. S. Cardoso, Jânio Monteiro, Roberto Lam, Valeria V. Krzhizhanovskaya, Michael H. Lees, Jack J. Dongarra, and Peter M.A. Sloot, editors, *Computational Science – ICCS 2019*, pages 126–140, Cham, 2019. Springer International Publishing.

[38] Azam Moosavi and Adrian Sandu. A state-space approach to analyze structural uncertainty in physical models. *Metrologia*, 2017.

[39] Azam Moosavi, Răzvan Ştefănescu, and Adrian Sandu. Multivariate predictions of local reduced-order-model errors and dimensions. *International Journal for Numerical Methods in Engineering*, 113(3):512–533, 2018.

[40] Hugh Morrison, Gregory Thompson, and V Tatarskii. Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one-and two-moment schemes. *Monthly Weather Review*, 137(3):991–1007, 2009.

[41] Kevin P Murphy. *Machine learning: A probabilistic perspective*. MIT press, 2012.

[42] Nasrin Nasrollahi, Amir AghaKouchak, Jialun Li, Xiaogang Gao, Kuolin Hsu, and Soroosh Sorooshian. Assessing the impacts of different WRF precipitation physics in hurricane simulations. *Weather and Forecasting*, 27(4):1003–1016, 2012.

[43] I Michael Navon, Xiaolei Zou, J Derber, and J Sela. Variational data assimilation with an adiabatic version of the NMC spectral model. *Monthly weather review*, 120(7):1433–1446, 1992.

[44] National Oceanic and Atmospheric Administration (NOAA). https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forcast-system-gfs.

[45] D. Orrell, L. Smith, J. Barkmeijer, and T.N. Palmer. Model error in weather forecasting. *Nonlinear Processes in Geophysics*, 8:357–371, 2001.

[46] T.N. Palmer, G.J. Shutts, R. Hagedorn, F.J. Doblas-Reyes, T. Jung, and M. Leutbecher. Representing model uncertainty in weather and climate prediction. *Annu. Rev. Earth Planet. Sci*, 33:163–93, 2005.

[47] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[48] J. Ross Quinlan. Induction of decision trees. *Machine learning*, 1(1):81–106, 1986.

[49] J Ross Quinlan. *C4. 5: programs for machine learning*. Elsevier, 2014.

[50] Vishwas Rao and Adrian Sandu. A posteriori error estimates for the solution of variational inverse problems. *SIAM/ASA Journal on Uncertainty Quantification*, 3(1):737–761, 2015.

[51] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning internal representations by error propagation. Technical report, DTIC Document, 1985.

[52] Mark R Segal. Machine learning benchmarks and random forest regression. *Center for Bioinformatics & Molecular Biostatistics*, 2004.

[53] Dong-Jun Seo and JP Breidenbach. Real-time correction of spatially nonuniform bias in radar rainfall data using rain gauge measurements. *Journal of Hydrometeorology*, 3(2):93–111, 2002.

[54] Carolin Strobl, Anne-Laure Boulesteix, Thomas Kneib, Thomas Augustin, and Achim Zeileis. Conditional variable importance for random forests. *BMC bioinformatics*, 9(1):307, 2008.

[55] Wei-Kuo Tao, Joanne Simpson, and Michael McCumber. An ice-water saturation adjustment. *Monthly Weather Review*, 117(1):231–235, 1989.

[56] Gregory Thompson, Paul R Field, Roy M Rasmussen, and William D Hall. Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. part ii: Implementation of a new snow parameterization. *Monthly Weather Review*, 136(12):5095–5115, 2008.

[57] Yannick Tr'emolet. Accounting for an imperfect model in 4D-Var. *Quarterly Journal of the Royal Meteorological Society*, 132(621):2483–2504, 2006.

[58] Yannick Trémolet. Model-error estimation in 4D-Var. *Quarterly Journal of the Royal Meteorological Society*, 133(626):1267–1280, 2007.

[59] Jiali Wang and Veerabhadra R Kotamarthi. Downscaling with a nested regional climate model in near-surface fields over the contiguous united states. *Journal of Geophysical Research: Atmospheres*, 119(14):8778–8797, 2014.

[60] Wei Wang and Nelson L Seaman. A comparison study of convective parameterization schemes in a mesoscale model. *Monthly Weather Review*, 125(2):252–278, 1997.

[61] Weather Research Forecast Model. https://www.mmm.ucar.edu/weather-research-and-forecasting-model.

[62] WRF Model Physics Options and References. http://www2.mmm.ucar.edu/wrf/users/phys_references.html.

[63] Dusanka Zupanski and Milija Zupanski. Model error estimation employing an ensemble data assimilation approach. *Monthly Weather Review*, 134(5):1337–1354, 2006.