



A Learning-Based Approach for Uncertainty Analysis in Numerical Weather Prediction Models

Azam Moosavi¹, Vishwas Rao^{2(✉)}, and Adrian Sandu³

¹ Biomedical Engineering Department,
Case Western Reserve University, Cleveland, USA
azmosavi@vt.edu

² Mathematics and Computer Science Division,
Argonne National Laboratory, Lemont, USA
vhebbur@anl.gov

³ Computational Science Laboratory, Department of Computer Science,
Virginia Tech, Blacksburg, USA
asandu7@vt.edu

Abstract. 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. We aim to address the following problems: (1) estimation of systematic model errors in output quantities of interest at future times and (2) identification of specific physical processes that contribute most to the forecast uncertainty in the quantity of interest under specified meteorological conditions. To address these problems, we employ simple machine learning algorithms and perform numerical experiments with Weather Research and Forecasting (WRF) model and the results show a reduction of forecast errors by an order of magnitude.

Keywords: Numerical weather prediction · Structural uncertainty · Model errors · Machine learning

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; such errors manifest themselves as missing dynamics in the model [11]. Examples of structural errors include the misrepresentation of sea-ice in the spring and fall, errors affecting the stratosphere above polar regions in winter [22], and 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, 15]. While traditional data assimilation reduces the uncertainty in the model state and model parameter values, however, no methodologies to reduce the structural model uncertainty are available to date.

In this study we consider the Weather Research and Forecasting (WRF) model [24], 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 [4, 14]. Among different atmospheric phenomena, the prediction of precipitation is extremely challenging and is obtained by solving the atmospheric dynamic and thermodynamic equations [14]. Model forecasts of precipitation are sensitive to physics options such as the microphysics, cumulus, long-wave, and short-wave radiation [5, 9, 14].

This paper demonstrates the potential of machine learning techniques to help solve two important problems related to the structural or physical uncertainty in numerical weather prediction models: (1) estimation of systematic model errors in output quantities of interest at future times, and the use of this information to improve the model forecasts, (2) 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 [6] a kernel-based regression method is developed as a forecasting approach with performance close to an ensemble Kalman filter (EnKF). Krasnopol et al. [8] employ an artificial neural network (ANN) technique for developing an ensemble stochastic convection parameterization for climate models.

This study focuses on the uncertainty in forecasts of cumulative precipitation caused by imperfect representations of the 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 the WRF prediction uncertainty in advance. 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 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 by supervised machine learning techniques, specifically, using ANN and Random Forests (RF) [13].

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 Sect. 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, for example, atmospheric model uncertainties are associated with subgrid modeling, boundary conditions, and forcings. All these modeling uncertainties are aggregated into a component that is generically called *model error* [7, 17, 18]. In the past decade considerable scientific effort has been spent in incorporating model errors and estimating their impact on the best estimate in both variational and statistical approaches [1, 20–22].

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

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

The state vector $\mathbf{x}_t \in \mathbb{R}^n$ contains the dynamic variables of the atmosphere such as temperature, pressure, precipitation, and tracer concentrations, 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 v_t , which are unknown. The atmosphere, as an abstract physical process, evolves in time as follows:

$$v_t = \mathcal{P}(v_{t-1}), \quad t = 1, \dots, T. \quad (1b)$$

The model state seeks to approximate the physical state:

$$\mathbf{x}_t \approx \psi(v_t), \quad t = 1, \dots, T, \quad (1c)$$

where the operator ψ maps the physical space onto the model space, for example, by sampling the continuous meteorological fields onto a finite-dimensional computational grid [11].

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 reality:

$$\psi(v_t) = \mathcal{M}(\psi(v_{t-1}), \Theta) + \boldsymbol{\delta}_t(v_t), \quad t = 1, \dots, T, \quad (2)$$

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

$$\mathbf{y}_t = h(v_t) + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \mathbf{R}_t), \quad t = 1, \dots, T. \quad (3)$$

Here h is the observation operator that maps the true state of atmosphere to the observation space, and the observation error $\boldsymbol{\epsilon}_t$ is assumed to be normally distributed.

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

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

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 (3) of the real system and the model predicted values of these observables (4) represent the model error in observation space:

$$\Delta_t = \mathbf{o}_t - \mathbf{y}_t \in \mathbb{R}^m, \quad t = 1, \dots, T. \quad (5)$$

For clarity, in what follows we make the following simplifying assumptions [11]:

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

These assumptions imply that the discretization errors are very small and that the main sources 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 assumed to be negligible.

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

$$v_t = \mathcal{M}(v_{t-1}, \Theta) + \delta_t(v_t), \quad t = 1, \dots, T, \quad (6a)$$

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

The model errors δ_t (2) are not fully known at any time t , since 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, Δ_τ for $\tau = t-1, t-2, \dots$, in order to estimate the model error δ_τ at future times $\tau = t, t+1, \dots$. This is achieved by unraveling the hidden information in the past Δ_τ values. Good estimates of the discrepancy δ_t , when available, could improve model predictions by applying the correction (6a) to model results:

$$\mathbf{v}_t \approx \mathbf{x}_t + \delta_t. \quad (7)$$

3 Approximating Model Errors Using Machine Learning

We propose a multivariate input-output learning model to predict the model errors δ , 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 $\Lambda \in \mathbb{R}^o$:

$$\phi(F) \approx \Lambda, \quad (8)$$

and approximate the function ϕ using machine learning.

In what follows we explain the problems we wish to address, function mapping ϕ , 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. We are interested in the uncertainty caused by the interactions 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. We define the following mapping:

$$\phi^{\text{error}}(\Theta, \mathbf{o}_\tau, \Delta_\tau, \mathbf{o}_t) \approx \Delta_t \quad \tau < t. \quad (9)$$

We use a machine learning algorithm to approximate the function ϕ^{error} . Using a supervised learning process, the learning model identifies the effect of physical packages, historical WRF forecast, historical model discrepancy, and the current WRF forecast on the available model discrepancy at the current time. After the model gets trained on the historical data, it yields an approximation to the mapping ϕ^{error} , denoted by $\hat{\phi}^{\text{error}}$. During the test phase the approximate mapping $\hat{\phi}^{\text{error}}$ is used to estimate the model discrepancy $\hat{\Delta}_{t+1}$ in advance. We emphasize that the model prediction (WRF forecast) at the time of interest $t+1$ (\mathbf{o}_{t+1}) is available, whereas the model discrepancy $\hat{\Delta}_{t+1}$ is an unknown quantity. In fact the run time of WRF is much smaller than the time interval between t and $t+1$; in other words, the time interval is large enough to run the WRF model and obtain the forecast for the next time window, estimate the model errors for the next time window and improve the model forecast by combining the model forecast and model errors. At the test time we predict the future model error as follows:

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

As explained in [11], the predicted error $\hat{\Delta}_{t+1}$ in the observation space can be used to estimate the error δ_{t+1} in the model space as follows:

$$\Delta_{t+1} \approx \mathbf{H}_t \cdot \delta_{t+1}, \quad \hat{\delta}_{t+1} \approx \mathbf{H}_t (\mathbf{H}_t^T \mathbf{H}_t)^{-1} \mathbf{H}_t^T \cdot \hat{\Delta}_{t+1}, \quad (10a)$$

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

$$\hat{\delta}_{t+1} \approx \text{cov}(\mathbf{x}_t, \mathbf{o}_t) (\text{cov}(\mathbf{o}_t, \mathbf{o}_t) + \mathbf{R}_t)^{-1} \hat{\Delta}_{t+1}, \quad (10b)$$

where \mathbf{R}_t is the covariance of observation errors.

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 interact with each other. 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 (for example time of the year, representation of sea-ice). The primary focus of this study is to accurately forecast the cumulative precipitation; therefore we seek to learn the impacts of all the physical packages that affect precipitation. To this end, we define the following mapping:

$$\phi^{\text{physics}}(\Delta_t) \approx \Theta, \quad (11)$$

which estimates the configuration Θ of the physical packages such that the WRF run generates a forecast with an error consistent with the prescribed level Δ_t (where Δ_t defined in Eq. (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 $\Delta_{\tau,i}^{\text{train}}$ for $\tau \leq t$ and $i \in \{\text{training data set}\}$. The output data is the corresponding physical combinations Θ that leads to the input error threshold.

To estimate the physics 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 Δ_t is split into a training part and a testing part. In the test phase we use the approximated function $\hat{\phi}^{\text{physics}}$ to estimate the physical process settings $\hat{\Theta}_j^1$ that are consistent with the observable errors $\Delta_{t,j}^{\{1\}}$. Here we select $\Delta_{t,j}^{\{1\}} = \Delta_{t,j}^{\text{test}}$ for each $j \in \{\text{test data set}\}$. Note that in this case, since we know what physics has been used for the current results, we can take $\hat{\Theta}_j^{\{1\}}$ to be the real parameter values $\Theta_j^{\{1\}}$ used to generate the test data. In general, $\Delta_{t,j}^{\{1\}}$ is chosen appropriately for a given application and the corresponding parameters are estimated.

Next, we reduce the desired forecast error level to $\Delta_{t,j}^{\{2\}} = \Delta_{t,j}^{\{1\}}/2$ and use the approximated function $\hat{\phi}^{\text{physics}}$ to estimate the physical process setting $\hat{\Theta}_j^{\{2\}}$ 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 $\hat{\Theta}_j^{\{2\}} - \hat{\Theta}_j^{\{1\}}$. Specifically, the number of times the setting of a physical process in $\hat{\Theta}_j^{\{2\}}$ is different from its setting in $\hat{\Theta}_j^{\{1\}}$ 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 to the model errors as estimated by the learning model.

3.3 Machine Learning Algorithms

To approximate the functions (9) and (11), we use regression machine learning methods. Choosing the right learning algorithm is challenging; it largely depends on the problem and the data available [2, 3, 10, 12]. Here, we use RF and ANN as our learning algorithms [13]. Both RF and ANN algorithms can handle nonlinearity in regression and classification. Given that the physical phenomena governing precipitation are highly nonlinear, and atmospheric dynamics is chaotic, we believe that RF and ANN approaches are well suited to capture the associated features. Although there are several other advanced machine learning algorithms that can be deployed here, we note that our aim here is to demonstrate the potential of machine learning approaches and hence the use of simple learning models such as RFs and ANN. Advanced techniques such as long short term memory (LSTM), convolutional neural networks (CNN), and gated recurrent units (GRU) are typically used for handling time series data and we defer such a study with these techniques to our future research. We describe the details regarding the training procedure, selection of the algorithm parameters, obtaining training data, and validation procedure in Sect. 4.

4 Numerical Experiments

We apply the proposed learning models to the WRF model [24] to (1) predict the bias in precipitation forecast caused by structural model errors, (2) predict the statistics associated with the precipitation errors, and (3) identify the specific physics packages that contribute most to precipitation forecast errors for given meteorological conditions.

4.1 WRF Model

In this study we use the non hydrostatic WRF model version 3.3. The simulation domain, shown in Fig. 1, 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 60 km [23]. The grid has 60 vertical levels to cover the troposphere and lower part of the stratosphere between the surface to approximately 20 km. 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 [16]. The stage IV estimates are available at an hourly temporal resolution over the continental United States. For experimental purposes, we use the stage IV NCEP analysis as a proxy for the true state of the atmosphere. The simulation window begins at 6AM UTC (Universal Time Coordinated) on May 1, 2017.

The model configuration parameters Θ represent various combinations of microphysics schemes, cumulus parameterizations, short-wave, and long-wave radiation schemes. Specifically, each process is represented by the schema values of each physical parameter it uses, as detailed in WRF model physics options

and references [25]. A total of 252 combinations of the four physical modules are used in the simulations. For each of the combinations, the effect of each physics combination on precipitation is investigated. The NCEP analysis grid points are 428×614 , while the WRF computational model have 60×73 grid points. To obtain the discrepancy between the WRF forecast and NCEP analysis, we linearly interpolate the analysis to transfer the physical variables onto the model grid. Figure 1 shows the NCEP analysis at 12PM on May 1, 2017 which is used as “true” (verification) state. Figure 2 shows the forecast at 12PM on May 1, 2017. For the initial conditions, we use the NCEP analysis at 6PM on May 1, 2017. The WRF forecast corresponding to the physics microphysics: Kessler, cumulus physics: Kain-Fritsch, long-wave radiation physics: Cam, short-wave radiation physics: Dudhia is illustrated in Fig. 2. Figures 5 and 6 shows contours of discrepancies at 12PM ($\Delta_{t=12\text{PM}}$) discussed in Eq. (5) for two physical combinations, which illustrates the effect that changing the physical schemes has on the forecast.

4.2 Experiments for Problem One: Predicting Pointwise Precipitation Forecast Errors over a Small Geographic Region

We demonstrate our learning algorithms to forecast precipitation in the state of Virginia on May 1, 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 Sect. 3.1, we learn the function ϕ^{error} of Eq. (9) using the training data from the previous forecast window (6AM to 12PM):

$$\phi^{\text{error}}(\Theta, \mathbf{o}_\tau, \Delta_\tau, \mathbf{o}_{t=12\text{PM}}) \approx \Delta_{t=12\text{PM}}, \quad 7\text{AM} \leq \tau < 12\text{PM}.$$

We use two learning algorithms to approximate the function ϕ^{error} . Specifically, 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 by using Scikit-learn, a machine learning library in Python [19]. For training purposes, we use the NCEP analysis of May 1, 2017 at 6AM as initial conditions for the WRF model. The forecast window is 6 h, and the WRF model forecast final simulation time is 12PM. The input features are as follows

- The physics combinations (Θ).
- The hourly WRF forecasts projected onto observation space \mathbf{o}_τ , $\text{AM} \leq \tau \leq 12\text{PM}$. The WRF state (\mathbf{x}_t) includes all model variables such as temperature, pressure, and precipitation. The observation operator extracts the precipitation portion of the WRF state vector, $\mathbf{o}_t \equiv \mathbf{x}_t^{\text{precipitation}}$. Accordingly, Δ_t is the discrepancy between WRF precipitation forecast \mathbf{o}_t and the observed precipitation \mathbf{y}_t .
- The observed discrepancies at past times (Δ_τ , $7\text{AM} \leq \tau < 12\text{PM}$).

The output variable is the discrepancy between the NCEP analysis and the WRF forecast at 12PM, that is, the observable discrepancies for the current forecast

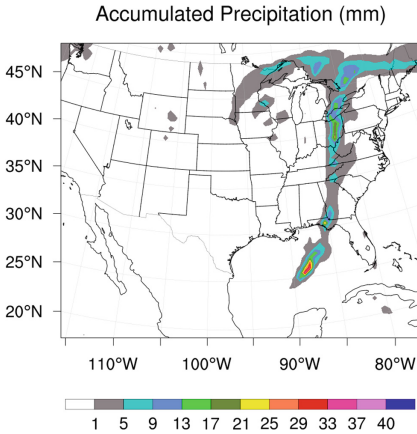


Fig. 1. NCEP analysis at 12PM provides a proxy for the true state of the atmosphere.

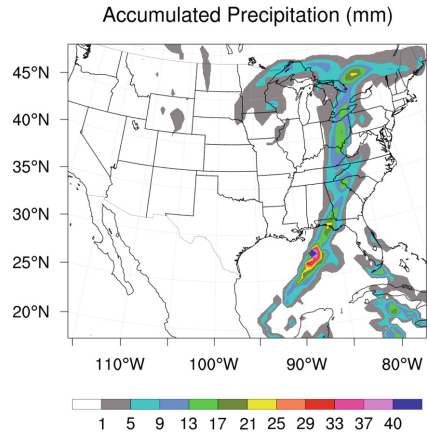


Fig. 2. WRF forecast at 12PM.

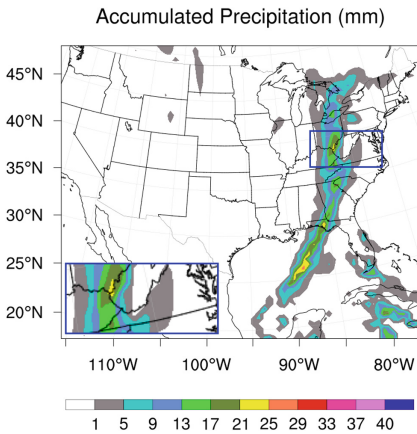


Fig. 3. Original WRF prediction at 6PM on May 1, 2017. Zoom-in panels show the predictions over Virginia.

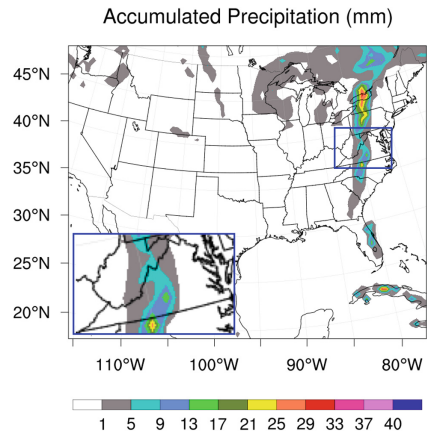


Fig. 4. NCEP analysis at 6PM on May 1, 2017. Zoom-in panels show the predictions over Virginia.

window ($\Delta t=12PM$). In fact, for each of the 252 different physical configurations, the WRF model forecast and 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.

Both ANN and RF are trained with these input-output combinations to obtain an approximation of the function ϕ^{error} , denoted by $\hat{\phi}^{error}$, during the

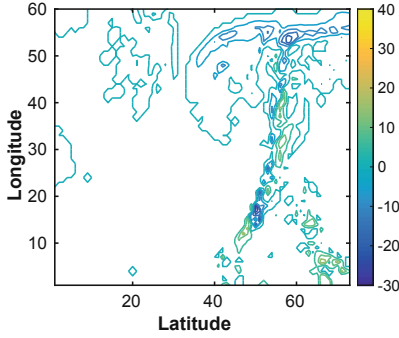


Fig. 5. Microphysics scheme: Kessler; cumulus physics: Kain-Fritsch; short wave radiation: Cam; long wave radiation: Dudhia

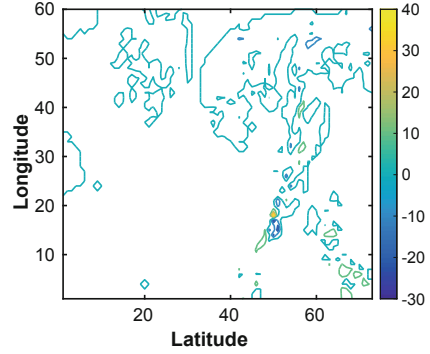


Fig. 6. micro-physics scheme: Lin; cumulus physics: Kain-Fritsch; short wave radiation: RRTM Mlawer; long wave radiation: Cam

training phase. In the testing phase we use the function $\hat{\phi}^{\text{error}}$ to predict the future forecast error $\hat{\Delta}_{t=6\text{PM}}$ given the combination of physical parameters as well as the WRF forecast at time 6PM as input features.

$$\hat{\Delta}_{t=6\text{PM}} \approx \hat{\phi}^{\text{error}}(\Theta, \mathbf{o}_\tau, \Delta_\tau, \mathbf{o}_{t=6\text{PM}}), \quad 1\text{PM} \leq \tau < 6\text{PM}.$$

To quantify the accuracy of the predicted error we calculate the root mean squared error (RMSE) between the true and predicted discrepancies at 6PM:

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

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

Table 1. Minimum, average, and maximum RMSE between the predicted $\hat{\Delta}_{t=6\text{PM}}$ and the true $\Delta_{t=6}$ over 252 physics combinations.

	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}

Figure 3 shows the WRF forecast for 6PM for the state of Virginia using the following physics packages (the physics options are given in parentheses): Microphysics (Kessler), cumulus-physics (Kain), short-wave radiation physics (Dudhia), and long-wave radiation physics (Janjic).

Table 2. Minimum and average of $\Delta_{t=6\text{PM}}$ for the original WRF forecast vs the improved forecast

	Minimum($\Delta_{t=6\text{PM}}$)	Average($\Delta_{t=6\text{PM}}$)
Original forecast	6.751×10^{-2}	5.025×10^{-1}
Improved forecast	2.134×10^{-4}	6.352×10^{-2}

Figure 4 shows the NCEP analysis at time 6PM, which is our proxy for the true state of the atmosphere. The discrepancy between the NCEP analysis and the raw WRF forecast is shown in the Fig. 7. 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 analysis is shown in the Fig. 8. The results show a considerable reduction of model errors when compared with the uncorrected forecast of Fig. 7. Table 2 shows the minimum and average of original model error vs the improved model errors.

4.3 Experiments for Problem Two: Identifying the Physical Processes that Contribute Most to the Forecast Uncertainty

The interaction of different physical processes greatly affects the 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:

$$\phi^{\text{physics}}(\Delta_{t=12\text{PM}}, \|\Delta_{t=12\text{PM}}\|_2) \approx \Theta.$$

Statistical characteristics include the mean, minimum, maximum, and variance of the field across all grid points over the continental United States. Note that this is slightly different from (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 $\Delta_{t=12\text{PM}}$ and $\|\Delta_{t=12\text{PM}}\|_2$.

In order to build the dataset, the WRF model is simulated for each of the 252 physical configurations, and the mismatches between the WRF forecasts and the NCEP analysis at the end of the current forecast window are obtained. The discrepancy between the NCEP analysis at 12PM and the WRF forecast at 12PM forms the observable discrepancy for the current forecast window $\Delta_{t=12\text{PM}}$. For

each of the 252 physical configurations, this process is repeated and statistical characteristics of the WRF forecast model error $\Delta_{t=12PM}$, and the norm of model error $\|\Delta_{t=12PM}\|_2$ 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 $\hat{\phi}^{\text{physics}}$. This function is applied to each of the 50 test samples $\Delta_{t=12PM}^{\text{test}}$ to obtain the predicted physical combinations $\hat{\Theta}_1$. To evaluate these predictions, we run the WRF model again with the $\hat{\Theta}_1$ physical setting and obtain the new forecast $\hat{\mathbf{o}}_{t=12PM}$ and the corresponding observable discrepancy $\hat{\Delta}_{t=12PM}^{\text{test}}$. The RMSE between the norm of actual observable discrepancies and the norm of predicted discrepancies is shown in Table 3. The small values of the difference demonstrates the performance of the learning algorithm.

Table 3. RMSE between estimated discrepancy using predicted physical combinations $\hat{\Delta}_{t=12PM}^{\text{test}}$ and the reference discrepancy $\Delta_{t=12PM}^{\text{test}}$.

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

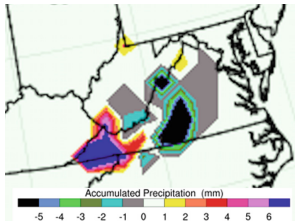


Fig. 7. Discrepancy between original WRF forecast and NCEP analysis

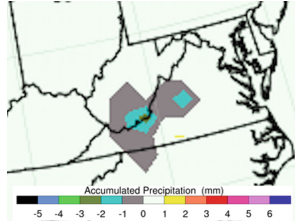


Fig. 8. Discrepancy between the corrected WRF forecast and the NCEP analysis

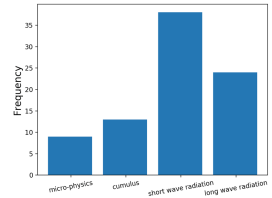


Fig. 9. Frequency of change in the physics with respect to change in the input data from $\Delta_{t=12PM}^{\text{test}}$ to $\Delta_{t=12PM}^{\text{test}}/2$. Each data set contains 50 data points, and we report the number of changes of each package.

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 $\Delta_{t=12\text{PM}}^{\text{test}}/2$ as inputs and obtain the predicted physical combinations $\hat{\Theta}_2$. The large variability in the predicted physical settings $\hat{\Theta}$ indicates that the WRF forecast error is sensitive to the corresponding physical packages. We count the number of times the predicted physics $\hat{\Theta}_2$ is different from $\hat{\Theta}_1$ when the input data spans the entire test data set.

The results shown in Fig. 9 indicate that microphysics 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 or incorrect physics.

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. 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 estimate the systematic model error in a quantity of interest at future times, and identify the physical processes that contribute most to the forecast uncertainty in a given quantity of interest under specified conditions.

Numerical experiments are performed with the WRF model using the NCEP analysis 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 illustrate the ability to estimate model errors, indicate the best model configuration, and choose the physical packages that influence WRF prediction accuracy the most.

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. The authors would like to thank Dr. Răzvan Ștefănescu for his valuable assistance and suggestions regarding WRF runs and the NCEP dataset.

References

1. Akella, S., Navon, I.M.: Different approaches to model error formulation in 4D-Var: a study with high-resolution advection schemes. *Tellus A* **61**(1), 112–128 (2009)
2. Asgari, E., Bastani, K.: The utility of Hierarchical Dirichlet Process for relationship detection of latent constructs. In: *Academy of Management Proceedings* (2017)
3. Attia, A., Moosavi, A., Sandu, A.: Cluster sampling filters for non-Gaussian data assimilation. arXiv preprint [arXiv:1607.03592](https://arxiv.org/abs/1607.03592) (2016)
4. Fovell, R.G.: Impact of microphysics on hurricane track and intensity forecasts. In: *Preprints, 7th WRF Users' Workshop, NCAR* (2006)
5. Fovell, R.G.: Influence of cloud-radiative feedback on tropical cyclone motion. In: *29th Conference on Hurricanes and Tropical Meteorology* (2010)
6. Gilbert, R.C., Richman, M.B., Trafalis, T.B., Leslie, L.M.: Machine learning methods for data assimilation. In: *Computational Intelligence in Architecturing Complex Engineering Systems*, pp. 105–112 (2010)
7. Glimm, J., Hou, S., Lee, Y., Sharp, D., Ye, K.: Sources of uncertainty and error in the simulation of flow in porous media. *Comput. Appl. Math.* **23**, 109–120 (2004)
8. Krasnopol'sky, V., Fox-Rabinovitz, M., Belochitski, A., Rasch, P.J., Blossey, P., Kogan, Y.: 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)
9. Lowrey, M.R.K., Yang, Z.L.: Assessing the capability of a regional-scale weather model to simulate extreme precipitation patterns and flooding in central Texas. *Weather Forecast.* **23**(6), 1102–1126 (2008)
10. Moosavi, A., Attia, A., Sandu, A.: A machine learning approach to adaptive covariance localization. arXiv preprint [arXiv:1801.00548](https://arxiv.org/abs/1801.00548) (2018)
11. Moosavi, A., Sandu, A.: A state-space approach to analyze structural uncertainty in physical models. *Metrologia* (2017). <http://iopscience.iop.org/10.1088/1681-7575/aa8f53>
12. Moosavi, A., Stefanescu, R., Sandu, A.: Multivariate predictions of local reduced-order-model errors and dimensions. arXiv preprint [arXiv:1701.03720](https://arxiv.org/abs/1701.03720) (2017)
13. Murphy, K.P.: *Machine Learning: A Probabilistic Perspective*. MIT Press, Cambridge (2012)
14. Nasrollahi, N., AghaKouchak, A., Li, J., Gao, X., Hsu, K., Sorooshian, S.: Assessing the impacts of different WRF precipitation physics in hurricane simulations. *Weather Forecast.* **27**(4), 1003–1016 (2012)
15. Navon, I.M., Zou, X., Derber, J., Sela, J.: Variational data assimilation with an adiabatic version of the NMC spectral model. *Monthly Weather Rev.* **120**(7), 1433–1446 (1992)
16. National Oceanic and Atmospheric Administration (NOAA). <https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forecast-system-gfs>
17. Orrell, D., Smith, L., Barkmeijer, J., Palmer, T.: Model error in weather forecasting. *Nonlinear Processes Geophys.* **8**, 357–371 (2001)
18. Palmer, T., Shutts, G., Hagedorn, R., Doblas-Reyes, F., Jung, T., Leutbecher, M.: Representing model uncertainty in weather and climate prediction. *Annu. Rev. Earth Planet. Sci.* **33**, 163–93 (2005)
19. Pedregosa, F., et al.: Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011)

20. Rao, V., Sandu, A.: A posteriori error estimates for the solution of variational inverse problems. *SIAM/ASA J. Uncertainty Quantification* **3**(1), 737–761 (2015)
21. Trémolet, Y.: Accounting for an imperfect model in 4D-Var. *Q. J. Roy. Meteorol. Soc.* **132**(621), 2483–2504 (2006)
22. Trémolet, Y.: Model-error estimation in 4D-Var. *Q. J. Roy. Meteorol. Soc.* **133**(626), 1267–1280 (2007)
23. Wang, J., Kotamarthi, V.R.: Downscaling with a nested regional climate model in near-surface fields over the contiguous united states. *J. Geophys. Res.: Atmos.* **119**(14), 8778–8797 (2014)
24. Weather Research Forecast Model. <https://www.mmm.ucar.edu/weather-research-and-forecasting-model>
25. WRF Model Physics Options and References. <http://www2.mmm.ucar.edu/wrf/users/phys.references.html>

Government License The submitted manuscript has been created by UChicago Argonne, LLC, Operator of Argonne National Laboratory (“Argonne”). Argonne, a U.S. Department of Energy Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357. The U.S. Government retains for itself, and others acting on its behalf, a paid-up nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government.