Visualization of Time-Varying Weather Ensembles Across Multiple Resolutions

Ayan Biswas, *Student Member, IEEE*, Guang Lin, Xiaotong Liu, and Han-Wei Shen

Abstract— Uncertainty quantification in climate ensembles is an important topic for the domain scientists, especially for decision making in the real-world scenarios. With powerful computers, simulations now produce time-varying and multi-resolution ensemble data sets. It is of extreme importance to understand the model sensitivity given the input parameters such that more computation power can be allocated to the parameters with higher influence on the output. Also, when ensemble data is produced at different resolutions, understanding the accuracy of different resolutions helps the total time required to produce a desired quality solution with improved storage and computation cost. In this work, we propose to tackle these non-trivial problems on the Weather Research and Forecasting (WRF) model output. We employ a moment independent sensitivity measure to quantify and analyze parameter sensitivity across spatial regions and time domain. A comparison of clustering structures across three resolutions enables the users to investigate the sensitivity variation over the spatial regions of the five input parameters. The temporal trend in the sensitivity values is explored via an MDS view linked with a line chart for interactive brushing. The spatial and temporal views are connected to provide a full exploration system for complete spatio-temporal sensitivity analysis. To analyze the accuracy across varying resolutions, we formulate a Bayesian approach to identify which regions are better predicted at which resolutions compared to the observed precipitation. This information is aggregated over the time domain and finally encoded in an output image through a custom color map that guides the domain experts towards an adaptive grid implementation given a cost model. Users can select and further analyze the spatial and temporal error patterns for multi-resolution accuracy analysis via brushing and linking on the produced image. In this work, we collaborate with a domain expert whose feedback shows the effectiveness of our proposed exploration work-flow.

Index Terms—Ensemble, time-varying, multi-resolution, sensitivity analysis

1 Introduction

Climate and weather simulation models are of prime importance to the scientists for analyzing and predicting future climate changes. For a reliable decision-making from these models, a robust prediction is needed for most likely scenarios (mean behavior) as well as low-probability but potentially hazardous events (outliers). To provide such robust predictions, full knowledge of the output, i.e. the full probability density function of the output is needed. Due to the lack of this ground truth, the output density function is generally approximated by running a set of ensemble simulations. A collection of techniques have been employed to generate ensemble simulations of climate model output such as using different simulation models, sampling the input parameter space, changing the initial condition of the simulations etc. Among these, perturbation of input parameters is most popular and well accepted [45] for generating ensembles of rainfall and other climate model outputs.

Despite the advancements in computation power and scientific capabilities, the true nature of how the different input parameters interact among themselves and with respect to the output, remains unknown. Due to the presence of non-linear interactions [42] among the physical processes, uncertainty quantification (UQ) of the model inputs and outputs is of prime importance. In climate modeling, physics parameterizations generally use conceptual or empirical relationships to approximate the impact of sub-grid processes on the resolved-scale dynamics and thermodynamics. Consequently, the parameterization schemes may contain some parameters that have no direct physical equivalence in nature and are subject to a certain degree of uncertainties. It remains unclear that such parametric uncertainty, parameter tuning and calibration may lead to different results in different regions and climate regimes. Thus a thorough analysis of the input parameters is required for good quality weather and climate prediction [4, 13, 41].

In this study, we use the Weather Research and Forecasting (WRF) regional climate model to examine the sensitivity and uncertainty on five convective parameters in the Kain-Fritsch (KF) convection parameterization scheme (CPS). The motivation of this study is to 1) investigate the spatial and temporal variation of parametric sensitivity and parameter estimation for the Kain-Fritsch convective scheme based on regional data under three different resolutions in the WRF model; 2) evaluate the trade-off between efficiency and accuracy among the ensemble of WRF model simulations under three different resolutions; 3) visualize the parametric sensitivity and uncertainty for the Kain-Fritsch convective scheme based on regional data under three different resolutions. To address these needs of the domain experts, we conducted a series of simulations to model the precipitation over the Southern Great Plains (SGP) area with the grid spacing of 12, 25, and 50 km (different spatial resolutions) and using two radiation schemes (RRTMG versus CAM) (different physical sub-models). The simulations produce a time-varying ensemble dataset across three resolutions along with their input parameterizations and with such high dimensional data set, visualization and analysis become non-trivial.

In this paper, we propose a work-flow for exploration of multi-resolution time-varying ensemble datasets that addresses the needs of the domain scientists. We attempt to answer two very specific questions posed by the domain experts: (1) How are the parametric sensitivity, parameter tuning and calibration processes sensitive to grid spacing, grid locations and temporal domain? (2) Are the optimal results transferable across spatial resolutions and how does the model accuracy change across resolutions?

For exploring the sensitivity of the input parameters against the output precipitation, a sensitivity measure named *delta* [3] is employed. Among the other existing sensitivity measures, this method is chosen since it is global, moment independent and popular in the operations research (OR) community. We apply this sensitivity measure across spatial locations and across time-domain to explore the intra and inter-

- *Ayan Biswas is with GRAVITY group, The Ohio State University. E-mail: ayanju04@gmail.com*
- *Guang Lin is with Purdue University. E-mail: guanglin@purdue.edu*
- *Xiaotong Liu is with GRAVITY group, The Ohio State University. E-mail: liu.1952@osu.edu*
- *Han-Wei Shen is with GRAVITY group, The Ohio State University. E-mail: shen.94@osu.edu*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx/

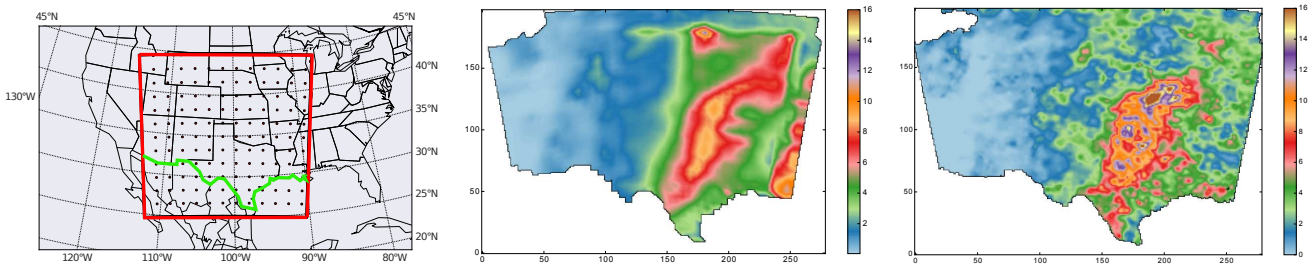


Fig. 1: Visualization of the ensemble dataset and the ground truth. (a) Southern Great Plains (SGP) regions for the ensemble data (marked in red). The green contour marks the United States boundary. Observed data is only available for the north of the green contour inside the red box. (b) Precipitation output at 50km resolution averaged over all the ensemble runs through the month of June, 2007. (c) Monthly averaged observed rainfall for the month of June, 2007.

resolution sensitivity trends. With sensitivity values distributed across spatial locations, a local clustering exposes the spatial trend in the sensitivity values. This clustering reveals the change of sensitivity with the model output and provides insight into which parameters are to be tuned finer for faster convergence of the simulations. To understand the temporal trend of the sensitivity values across resolutions, we apply a dimensionality reduction approach to create a two-dimensional layout along with a line chart view and allow the users to interactively explore the entire data domain by connecting it with the spatial view. To explore the efficiency and accuracy of the three resolutions, we take up a non-parametric kernel density estimation approach to analyze the ensemble precipitation outputs. Given the observed precipitation data, we employ Bayesian likelihood analysis to identify which resolution has better prediction accuracy at different spatial locations. After computing these local best predictors, the time-varying data is aggregated into one final image using custom color-maps that guides the future simulations towards an adaptive grid implementation. For detailed exploration of accuracy trends of the multi-resolution dataset, we design an interface with brushing and linking of the temporal and spatial domains. Our designed exploration tool uses domain experts' inputs and helps them understand the multi-resolution temporal ensemble dataset. We work closely with a domain expert at each stage of the pipeline to verify and answer his domain specific needs. The domain expert's feedback shows that our framework is effective in providing detailed insight into such a high-dimensional dataset.

Our contributions in this work are three-fold:

1. For ensemble data with input parameter settings, we introduce sensitivity analysis across spatial and time-domain for effective data visualization and exploration.

2. For temporal multi-resolution ensemble data with ground truth, we propose a new accuracy estimation model with uncertainty quantification.

3. We propose a complete work-flow for exploring time-varying multi-resolution ensemble data sets that works closely with the domain expert at all stages of the pipeline. Our framework puts the domain expert in the exploration loop and continually refines the output based on the needs of the domain expert.

2 Related Works

Our work focuses on detailed visual exploration of parameter sensitivity and accuracy analysis of the multi-resolution temporal ensemble datasets. Here we briefly describe some of the notable related works from the well researched fields of uncertainty and ensemble visualization and sensitivity analysis.

Uncertainty and Ensemble Visualization : Visualization of ensemble datasets is an important subtopic under the techniques of uncertainty visualization. The general uncertainty visualization methods have been the focus of research for a long time now [27]. Johnson and Sanderson [21] discussed the importance of the topic and provided a detailed discussion on the existing techniques of 2D and 3D uncertainty visualization including uncertainty in surfaces. Uncertainty for the isosurfaces has been a topic of extended research in the recent past. Probabilistic isosurface extraction algorithms have been proposed by Pöthkow et al. [29, 30] and Pfaffelmoser et al. [28]. The authors proposed different approaches depending on whether the correlation among the grid cells is considered or not. Schlegel et al. [37] recommended the use of Gaussian Process Regression for high fidelity interpolation of the uncertain data in the presence of Gaussian noise. For a discussion on more recent uncertainty visualization techniques, the readers are referred to the survey work by Potter et al. [31].

Simulations generating ensemble outputs are the main source of uncertain datasets. Use of ensemble datasets is quite common for the weather analysts and visual exploration of the uncertainty related to these ensemble datasets has been a popular topic of research. A framework called Ensemble-Vis was created by Potter et al. [32] for weather forecasting and climate modeling where they combined multiple ensemble visualization techniques. Sanyal et al. [36] proposed Noodles as an enhancement to the existing Spaghetti plots for understanding the meteorological datasets. They used circular glyphs and confidence ribbons to depict the Euclidean spread of isocontour lines effectively. Recently, Hao et al. [14] proposed a kd-tree based exploration technique for temporal particle ensemble datasets. Another visual verification system was proposed by Bock et al. [2] to explore the Coronal Mass Ejection phenomenon. Demir et al. [8] extended the bar and line charts to design Multi-charts that facilitated the comparative exploration of multiple volume datasets. They used bidirectional brushing and linking to facilitate user interaction and query visualization. For providing a comparative visualization of 2D temporal ensemble datasets, Höllt et al. [15, 17] introduced time-series glyphs. Other notable comparative visual analysis methods for vector ensembles were proposed by Guo et al. [11] and Jarema et al. [20]. Summarization methods for ensemble space curves were discussed in the works of Whitaker et al. [43], Mirzargar et al. [25] and Ferstl et al. [9]. Bensema et al. [1] introduced modality-based classification for the high variance regions of ensemble datasets to facilitate a detailed understanding of the ensemble distributions. Despite of these existing works, a thorough sensitivity study and accuracy analysis to connect both the input and output of ensemble datasets are missing. In our work, we design a visual analysis workflow that enables the domain experts for efficient exploration of sensitivity and accuracy of multi-resolution ensemble data sets.

Sensitivity Analysis : The goal of sensitivity analysis is to identify the most influential parameters from a set of uncertain input parameters for a simulation. This sensitivity study helps the scientists to understand and verify their simulation model, prioritize the parameters for more effective sampling of parameter space, and even simplify the simulation model. The existing measures for sensitivity analysis can be broadly classified into two groups: local and global [19]. The local methods aim at observing the change in the output when the input parameters go through small perturbations. These methods generally compute partial derivatives at important feature points e.g. around the

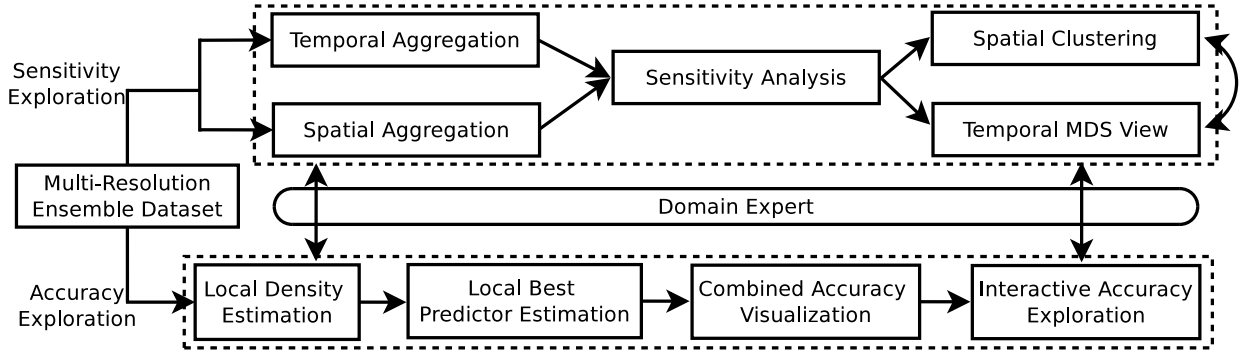


Fig. 2: A schematic overview of our algorithm.

mean. The local methods were popular in the earlier phase of sensitivity analysis works [5, 18, 22] and also have been applied for hydrological modeling recently [6]. Compared to the local techniques, global methods compute the sensitivity by varying the parameters throughout the possible distribution range [38] and are often preferred over the local counterparts [19]. A variance-based global sensitivity quantification method was proposed by Sobol [40]. Saltelli et al. [35] proposed a Fourier amplitude sensitivity test (FAST) for finding the total influence of individual parameters on the output. A moment-independent sensitivity measure was given by Borgonovo [3] that computed the sensitivity of parameters by observing the change in the output distribution. Among other popular methods, the derivative based global sensitivity measure (DGSM) by Sobol and Kucherenko [39], incremental ratio based input parameter screening method by Morris [26] are the most popular.

3 Background and Overview

3.1 Motivation

Along with the accuracy of the model output, the input parameter sensitivity study has historically been of prime importance for weather ensembles, and it continues to be an active research area [12, 16, 44, 46]. Weather and climate simulation models often include multiple input parameters ($\approx$ 15 to 30) and there generally exists a non-linear relationship among the inputs and outputs in these models. For such complex models, a lot of computational power and resources are expended in high-dimensional input parameter space sampling for generating good quality ensemble outputs. Consequently, the climate modelers employ parameter sensitivity analysis to rank parameters [12, 44, 46] according to their influence. This enables the scientists to calibrate the model by tuning the most influential parameters of the system or perform experiments to better characterize those parameters. When the simulations are used to generate time-varying data at different spatial locations, this sensitivity information of the parameters can change over the spatial-temporal domain. Thus, a thorough exploration of how the sensitivity values of the model parameters change as a function of spatial location and over the temporal domain, is of great importance to the domain experts.

Exploration of accuracy and parameter sensitivity becomes harder when the ensemble simulation output is generated for varying grid resolutions. Domain experts investigate multi-resolution datasets from simulation models since datasets of varying resolutions can capture different scales of physical phenomena. As different resolutions have different computational costs, domain experts need to understand the accuracy vs cost trade-offs for running future simulations. In climate simulations, as the model performance is sensitive to other physical parameterizations which also depend on model resolution, it is not guaranteed that higher resolution models always perform better than lower resolution models. For instance, it is observed that in some ensembles, precipitation results using 50km spacing grid can be more accurate than that using higher resolution 12km spacing grid. Hence it is critical to develop an effective data visualization tool to analyze simulation models for identifying the advantages and limitations of generating data at varying resolutions. The existing data analytics software, such as Paraview, VisIt, Tecplot, etc. do not integrate the multi-resolution aspect of the temporal ensembles for exploration of accuracy and sensitivity across spatial location and time-domain. Our proposed solution keeps the domain expert in the loop for a detailed and effective visualization-guided exploration of the entire dataset.

Table 1: Description and range of the input parameters for the WRF model.

Parameter	Description	Range	Default
Pe	Coefficient related to entrainment mass flux rate	[-1,1]	0
Pd	Coefficient related to downdraft mass flux rate	[-1,1]	0
Pt	Maximum turbulent kinetic energy	[3,12]	5
Ph	Starting height of downdraft above updraft source layer	[50,350]	150
Pc	Average consumption time of convective available potential energy	[900,2700]	2700

3.2 Data Description

WRF regional climate model is widely accepted mesoscale numerical weather prediction system and this model uses Kain-Fritsch (KF) convection parameterization scheme (CPS) to simulate convection. A set of five parameters is used in the simulations to generate the precipitation output. In particular, the five uncertain convective parameters (as input parameters to the ensemble WRF model simulations) are: Pd, coefficient related to downdraft mass flux rate, Pe, coefficient related to entrainment mass flux rate, Pt, maximum turbulent kinetic energy, Ph, starting height of downdraft above updraft source layer, and Pc, average consumption time of convective available potential energy. The range of values and the default values used for each of the parameters are listed in Table 1. The model parameters were calibrated using simulated stochastic approximation annealing (SSAA) calibration algorithm. The SSAA algorithm used for global calibration is based on the Very Fast Simulated Annealing (VFSA) algorithm [23]. Using this approach, the sampling range of each parameter narrows at every stage of the convergence process. We also calibrate the CPS by constraining simulated precipitation with the observations.

The observation dataset consists of observed precipitation in June 2007 over the Southern Great Plains (SGP) area (latitude $\times$ longitude) $[25.0, 44.0] \times [-112.0, -90.0]$. The SGP coverage area is shown in Figure 1a and the monthly averaged observed precipitation is depicted in Figure 1c. The computer simulation dataset consists of 150 ensemble simulations at different sets of 5 uncertain convective parameter values selected via Latin Hypercube Sampling [24] over the SGP area with the grid spacing of 12, 25, and 50 km. This generates 150 ensemble simulated output precipitation at three different resolutions over the 30 days. The regions outside United States of America and ocean regions are marked with a large positive value in the ensemble dataset and are discarded from the analysis. Monthly average of the mean ensemble output for 50km resolution is shown in Figure 1b.

3.3 Problem Statement

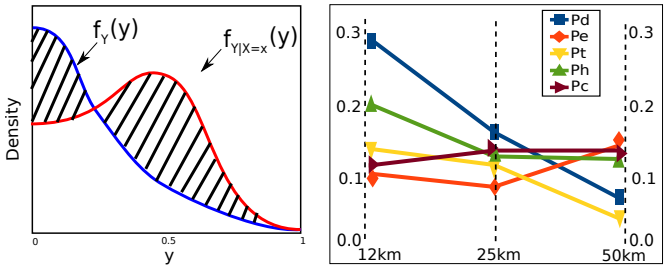
In the context of the WRF climate model, domain experts have a sequence of domain specific questions that we try to answer in this work. The specific questions that we address in this work include the following:

1. Given the set of input parameters and the ensemble output, which parameters are most important in deciding the output precipitation?
2. How the sensitivity of the individual parameters change as a function of spatial location, time step and across different resolutions?
3. Do the sensitivity values of the different input parameters remain the same over different spatial locations and time steps for different resolutions? Or, how the relative importance of the parameters change?
4. How the different data resolutions are related to each other in terms of their input parameter sensitivities? Do the sensitivities change when the ensemble output is generated at varying resolutions?
5. In the presence of the ground truth, how does the prediction accuracy of the ensemble runs change for different resolutions? Does the higher resolution always have higher accuracy?
6. How to effectively explore the regions with varying prediction accuracy of a multi-resolution temporal ensemble dataset? How to quickly identify the regions with high prediction error for different resolutions?

3.4 Visualization System Overview

Our work is focused on visualization-based exploration of multi-resolution temporal ensemble datasets. Our system design addresses the needs of the domain experts, and the visualization components used in this system are targeted towards answering the previously mentioned queries of the domain experts. As shown in Figure 2, our system has two modes of visual exploration: sensitivity exploration mode and accuracy analysis mode. In the sensitivity exploration mode, sensitivity visualization can be performed across resolutions by temporal and/or spatial aggregation of the input dataset. In spatial aggregation, users intend to understand the time-varying sensitivities of the input parameters, and to explore the similarity of these sensitivity values across resolutions. After performing sensitivity analysis using the δ method [3] for the five input parameters, users are presented with a multi-dimensional scaling (MDS) projection view along with a connected line chart view. This view helps users understand the relationship of the different resolutions and facilitates further user interaction for more detailed exploration. From the temporal view, users can select different days and explore the similarity in sensitivities of the input parameters as a function of spatial locations. Users can spatially explore the sensitivity information by spatial clustering of the sensitivities. For understanding the multi-resolution sensitivity, the clustering agreement is provided to the user using a custom color-map. Using this connected temporal and spatial view, users can effectively explore the sensitivity relationships by interacting with the system components.

For accuracy exploration, the observed precipitation information is utilized to provide a comparative view of the ensemble outputs at different resolutions. Since every spatial location has a set of ensemble precipitation output for every resolution, initially kernel density estimation is applied to approximate the density functions. Since different resolutions predict the observed precipitation differently, their accuracy is estimated from their corresponding kernel density estimates. A Bayes rule based method is utilized for this likelihood computation and one of the resolutions is identified as a local best predictor at a spatial location. Since the local predictors can be different over the time domain for a given location, this temporal variation is aggregated to generate an image for combined accuracy visualization. Users can interactively brush and select regions from this image to get information regarding the temporal error on a connected window showing the time-varying error trends. To further explore accuracy on user selected days, spatial error distribution is presented for all the resolutions to facilitate accuracy comparison. An error histogram based selection tool is linked to this multi-resolution view to quickly locate regions of high error. This interactive mode of exploration provides a detailed analysis of accuracy trade-offs across the resolutions.



(a) Sensitivity analysis measure δ computes the shift in the distribution of output variable Y when one parameter is fixed (image courtesy [3]).

(b) Sensitivity analysis measure δ computed at three resolutions of the spatio-temporal averaged data for the five input parameters.

Fig. 3: Schematic explanation of the δ method and its application on the aggregated data.

4 Sensitivity Analysis of Input Parameters

4.1 Global Sensitivity Measure

Sensitivity analysis is an important topic in operations research (OR) where the analysts strive to quantify the relationship between the input and output variables of their models while taking decisions. As discussed in Section 2, the sensitivity analysis measures can be broadly classified into two classes: local and global. In general, a good sensitivity analysis measure is expected to be "global, quantitative, and model free" [34]. Here, "model free" refers to the fact that the measure should not depend on any assumptions regarding the relationship among the inputs and outputs of the model. For the WRF model and its input parameters, a local sensitivity analysis was previously conducted [44] assuming the input parameters were uncorrelated. In this work, we use a sensitivity measure named δ that is both "global" and "model free". The δ method was proposed by Borgonovo [3]. Let us assume that the model under consideration is $g(\vec{X}) = Y$, where $\vec{X} = \{X_1, X_2, ..., X_n\} \in R^n$, the n uncertain input parameters, and Y is the model output. If $f(\cdot)$ represents the density function, then $f_Y(y)$ is the output density function when all the input parameters are allowed to vary unconditionally. Now, if the input parameter $X_i = x_i$, then the shift in the output distribution is given by the total area calculated as:

$$s(X_i, Y) = \int |f_Y(y) - f_{Y|X_i=x_i}(y)| dy. \tag{1}$$

For the ith input parameter, δ_i is then defined as :

$$\begin{aligned} \delta_i(Y) &= \frac{1}{2} E_{X_i}[s(X_i, Y)] \\ &= \frac{1}{2} \int f_{X_i}(x_i) \int |f_Y(y) - f_{Y|X_i=x_i}(y)| dy dx_i. \end{aligned} \tag{2}$$

Thus a higher value of δ_i denotes that the expected shift in the density function of the output is higher due to the input X_i, establishing the output to be more sensitive to X_i. Conversely, if the input parameter X_i has no influence on the output Y, there is no change in the density functions $f_Y(y)$ and $f_{Y|X_i=x_i}(y), \forall x_i \in X_i$. This results in $\delta = 0$ as $s(X_i) = 0$, since $f_Y(y) - f_{Y|X_i=x_i}(y) = 0, \forall x_i \in X_i$. The concept of δ measure is schematically presented in the Figure 3a. This global importance indicator is moment independent since it captures the change in output distribution rather than the change in the variance. As explained by Borgonovo [3], this measure has the property: $0 \leq \delta_i \leq 1$. We use

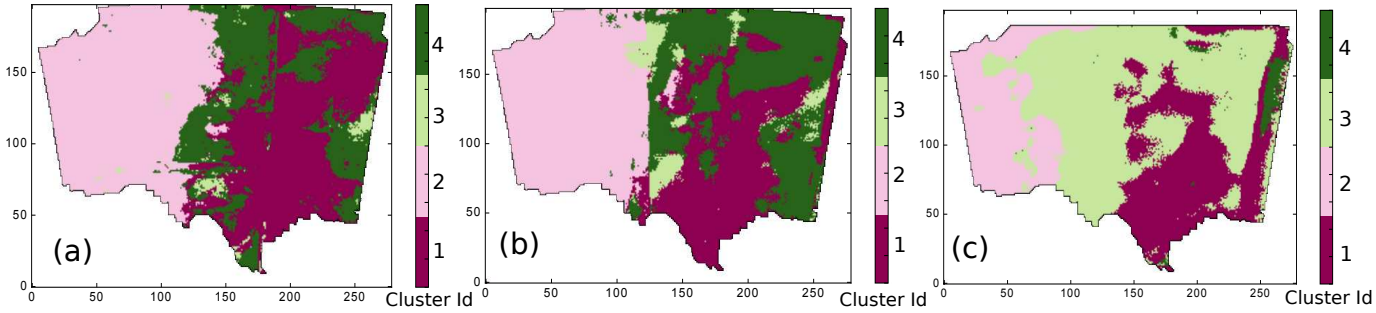


Fig. 4: Sensitivity-based clustering for three resolutions (a) 12km, (b) 25km, and (c) 50km.

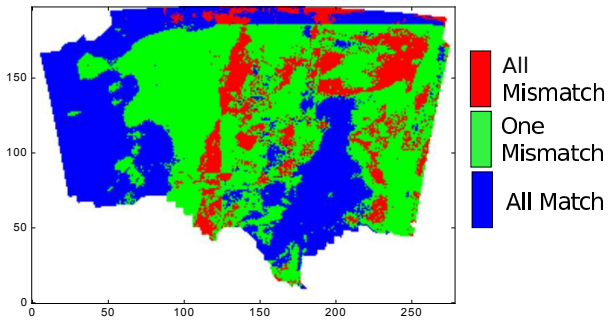


Fig. 5: The agreement among the three resolutions for the clustering presented in Figure 4.

this sensitivity measure for exploration of the weather ensembles as described in the following sections.

4.2 Exploration of Parameter Sensitivity Across Resolutions

Given the WRF model generated precipitation dataset, the domain experts want to investigate the sensitivity of the input parameters and how the sensitivity values change when the model runs at different spatial locations and at different days across the three resolutions. The domain experts can choose to explore the sensitivities at different levels of granularity: the finest level being observing the sensitivity of the input parameters at every grid location for every day, the coarsest level is to observe how the input parameters influence the monthly aggregated precipitation over the whole spatial domain. Initially, to provide a high level view of the sensitivity of the parameters, we first aggregate the dataset to find the most influential parameters and present that to the domain experts. We compute the spatio-temporal average of the precipitation values for each resolution. Next the δ method is applied and the results are shown in Figure 3b. This shows that Pd is the most important at 12km and 25km but Pe becomes the most important at 50km resolution. Starting from this information, to further understand how the parameter sensitivity changes spatio-temporally, we design two ways for the exploration of the sensitivity values: a) spatial exploration and b) temporal exploration. In the following sections, the time-varying ensemble precipitation output is denoted as $Y(t,\vec{v},r)$ where t represents days and $\vec{v} \in \{x,y\}$ represents spatial locations, and $r \in \{h,m,l\}$ denotes the different resolutions where h,m,l are abbreviations for *high*, *mid* and *low* respectively.

4.2.1 Spatial Exploration of Parameter Sensitivity

Understanding sensitivity as a function of the spatial locations provides an important insight into the selection of parameters based on locations. Moreover, when simulation models are used to generate data at varying resolutions, the sensitivities of parameters can change across resolutions. Given a period of time, understanding how location-based sensitivity values are comparable across resolutions is of great interest to the domain experts. In addition, weather analysts and domain experts need to apply averaging and smoothing at various levels temporally and spatially. Adhering to this need, after understanding the high level importance of the input parameters as presented in Figure 3b, we investigate the parameters into a finer detailed level by temporally aggregating the ensemble dataset over a user specified θ days. The ensemble output $Y(t,\vec{v},r)$ then reduces to $Y_\theta(\vec{v},r)$ as:

$$Y_\theta(\vec{v},r) = \frac{\sum_{t=1}^{\theta} w_t * Y(t,\vec{v},r)}{\sum_{t=1}^{\theta} w_t}. \quad (3)$$

Here t represents days and $\vec{v} \in \{x,y\}$ represents spatial locations, and $r \in \{h,m,l\}$ denotes the different resolutions. The aggregation weights w_t's are domain specific and dependent on the domain experts choice. The equal weighted w_t's can be used for initial data exploration purposes. Next, for a given resolution r, the importance measure δ is computed at every spatial location $\vec{v}$ for all the input parameters that produces sensitivity value tuples $\vec{\delta}$ at every spatial location, i.e., at every location $\vec{v}$ we have:

$$\vec{\delta}(Y_\theta(\vec{v},r)) \text{ with } \vec{\delta} = \{\delta_i\},\ i = \{1,...,5\}, \quad (4)$$

where i represents the different input parameters. After this stage, every spatial location now contains a five dimensional vector describing the sensitivity information of the input five parameters. To understand how these spatial locations are similar to each other according to their sensitivity vectors in a local region, a spatial clustering is applied. To achieve this spatial clustering, we use the k-means algorithm which is popular for its runtime and quality of clustering trade-off. Using this method, the feature vector $\vec{fv}$ consists of the sensitivity values of each location and the location information, i.e., for a given location $\vec{v} = (x,y)$ if the sensitivity tuple $\vec{\delta}$ is computed by following Equation 4, then we get $\vec{fv}$ as :

$$\vec{fv} = \{\vec{\delta}, w_{sp} * \vec{v}\}, \text{ with } 0 < w_{sp} < 1. \quad (5)$$

The spatial weight w_{sp} is a trade-off between how much the clustering is influenced by the spatial locations compared to the sensitivity values. If $w_{sp} = 0$, then this produces clustering of sensitivity values that are not influenced by the spatial locations.

Since our goal in this work is to compare the clustering results across resolutions, instead of computing clustering results for every resolution separately, we incorporate the information from all the resolutions together and then generate the clustering results. We concatenate the N data points from each of the three resolutions to form an input matrix M of dimension $(3N \times \left|\vec{fv}\right|)$. Since in our case $\left|\vec{fv}\right| = 7$, dimension of M is $(3N \times 7)$. Given k clusters as the input parameter to the k-means algorithm, now the individual spatial locations at each resolution can be assigned a color based on the output cluster id. This yields three images for the three resolutions that can be directly compared visually to understand how the sensitivity values agree both spatially and across resolutions. The clustering results of 12km, 25km and 50km resolutions are presented in Figure 4a-c where number of clusters $k = 4$ with $w_j = 0.009$ as the spatial weight factor and $\theta = 30$ representing a monthly average. It is visually apparent

that the simulations at 12km and 25km resolutions show quite similar clustering of the input parameter sensitivities. In comparison, the simulation run at the lowest resolution 50km, is much different from the other two. To convey more information through the visualization, the users can select a cluster and analyze the sensitivity of those regions. This helps the domain experts to observe how parameter sensitivities change across regions and resolutions to understand their simulations further.

Finally we provide a summarized clustering information to depict the agreement among the three resolutions. Each location $\vec{v}$ now contains three cluster ids $\{id_h, id_m, id_l\}$ originating from three resolutions. We present a combined visualization of the three clustering images from three resolutions to depict the agreement of these resolutions in an integrated view. Given $\vec{id} = \{id_h, id_m, id_l\}$ at every spatial location (x,y), a color conversion function $Cmap(\vec{id})$ is defined as:

$$Cmap(\vec{id}) = \begin{cases} blue, & \text{if } id_h = id_m = id_l \\ red, & \text{if } id_h \neq id_m \neq id_l \\ green, & \text{otherwise} \end{cases}$$

The agreement of the clustering information provided in Figure 4a-c is shown in Figure 4d. The blue color shows the region where all the three resolutions show similar clustering, red color shows the regions where all the resolutions mismatch and green shows where exactly two of the three resolutions match but the other one does not. This figure shows that in the regions where all the three resolutions predict high rain or very low rain, the sensitivities match. But for the regions where there are some variations in the predicted output precipitation, the sensitivity values also differ. This exploration motivates the users to further investigate their simulation parameters.

4.2.2 Temporal Exploration of Parameter Sensitivity

After exploring the sensitivity of input parameters spatially, our next goal is to understand the parameter importance in time-domain for all the three resolutions. The domain experts need to understand the sensitivity of the parameters on a daily basis or over a specific period of time. While making predictions regarding the precipitation of a given period, understanding which input parameters have the most influence is of prime importance. Understanding the progression of sensitivity values over time leads to a better optimization of parameter space sampling and faster convergence with better predictions. To address this domain need, in our design we assume an aggregation period of α days and we get the time-varying data $Y_\alpha(t,r)$ by computing the spatio-temporal average behavior of the ensemble precipitation dataset as:

$$Y_\alpha(t,r) = \frac{\sum_{k=t}^{t+\alpha} w_k * \sum_{\vec{v}} Y(k,\vec{v},r)}{\sum_{k=t}^{t+\alpha} w_i}, \; t = 1,...,D-\alpha+1. \qquad (6)$$

Here t represents time, r represents a given resolution and D is a constant representing the total number of days for which the simulation is run. The aggregation weights w_k's are again dependent on the simulation model. It can be readily observed that the Equation 6 will reduce to daily temporal analysis when $\alpha = 0$. When $\alpha = D$, it will reduce to the monthly spatio-temporal aggregation. For any other choices of α where $0 < \alpha < D$, $Y_\alpha(t,r)$ is the moving average of $\sum_{\vec{v}} Y(t,\vec{v},r)$ in temporal dimension. Now the importance measure δ is computed at $t, \forall t = 1,...,D-\alpha+1$ that produces $\vec{\delta}$ as:

$$\vec{\delta}(Y_\alpha(t,r)) \text{ where } \vec{\delta} = \{\delta_i\},\ i = \{1,...,5\}. \qquad (7)$$

For a systematic exploration of temporal sensitivity, we design a visualization component that depicts the information contained in $\delta(Y_\alpha(t,r))$. Since each day is now represented by a five dimensional feature vector $\{\delta_i\}$, and there are three different resolutions, a data transformation is required when the users want to explore the similarities of different resolutions over $D-\alpha+1$ days. To meet this need, we first apply dimensionality reduction technique to project the five-dimensional data points to two-dimensions by preserving the higher

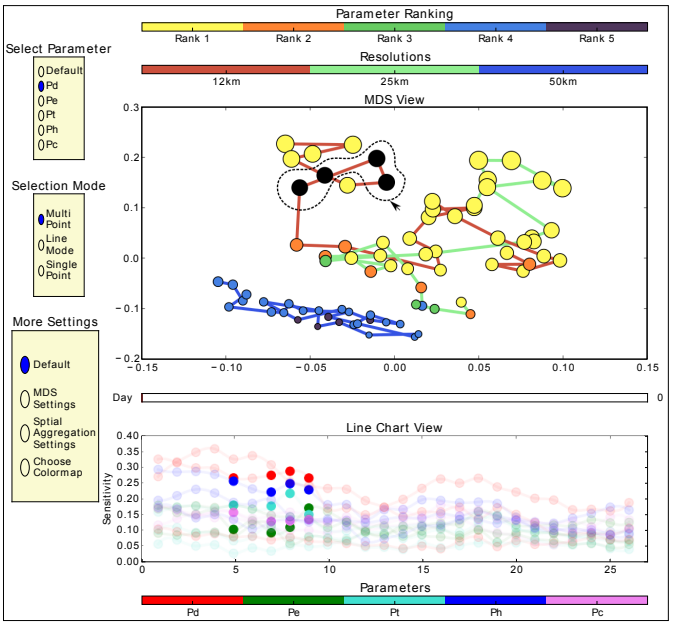


Fig. 6: Time-varying sensitivity visualization using our proposed technique. Top: the MDS plot shows the similarity among the different days across three resolutions where each point is a five dimensional vector of sensitivity values of the five parameters. Bottom: Line chart shows the detailed sensitivity values for the user selected days.

dimensional distances among the points as much as possible. Since each day is represented by a five dimensional feature vector for each of the three resolutions, there are total $3 \times (D-\alpha+1)$ such high dimensional points. In this work, we employ the popular multi-dimensional scaling (MDS) technique [7] to reduce the dimensionality of our input set of data points from five dimensions to two dimensions. MDS is popular since it is a non-linear dimensionality reduction technique and also it preserves the high dimensional distances in the projected lower dimension.

Application of MDS generates a scatter plot in two-dimensional space for the different days of the three resolutions as shown in Figure 6. In this figure, we have used the time aggregation period $\alpha = 5$. The points from the same resolution are connected by straight line segments where the color of the line segments designate the particular resolution (red for 12km, green for 25km, blue for 50km). From this MDS plot, it can be readily inferred that the lowest resolution behaves very differently compared to the other two resolutions when all the parameter sensitivity values are considered together. To show the progression of time, a "Day" slider tool-bar is used that can be tuned by the users to highlight the corresponding day in the MDS plot from all the three resolutions. From this MDS view, users can also compare the sensitivity values of the different days within and across resolutions. Since domain experts are interested in observing which parameter is the most important and how much important, we use the size and color of the MDS plots to convey these two pieces of information. For a selected parameter for each day and resolution, its rank is computed in comparison to the other parameters and it is color-mapped in the scatter plot according to a predefined discrete color-map. The absolute sensitivity score of the selected parameter is used to modulate the size of the scatter plot points. Using this design, more important parameters will be highlighted in larger size and brighter colors for the different days and resolutions. In Figure 6, parameter Pd is selected to demonstrate this design. As can be seen, Pd is very important in the 12km and 25km in deciding the precipitation, but at the lowest resolution 50km, Pd becomes less important.

After observing the high level view of the temporal trend, users can opt for detailed analysis of the individual days. Using a lasso-based selection tool, users can select days from different resolutions and simultaneously view them in a connected line chart. This line chart con-

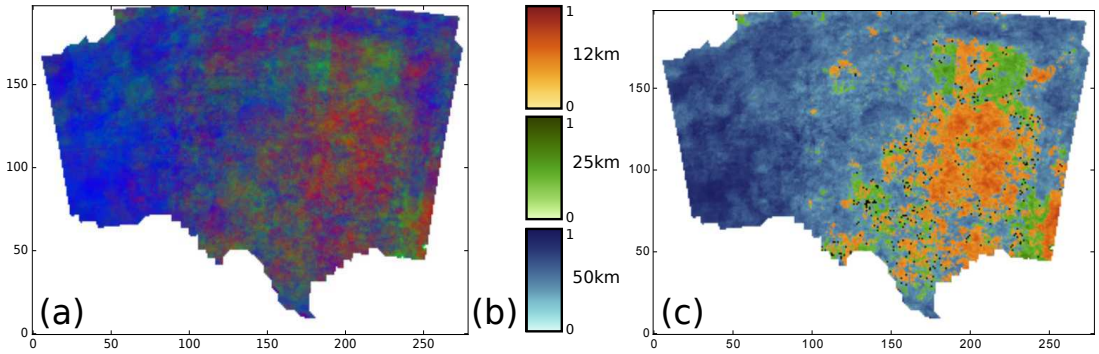


Fig. 7: Combined resolution map for accuracy exploration. (a) Simple rgb color-mapped visualization is ineffective for understanding the dataset. (b) Our proposed method of first classifying and then mapping to color is more useful for exploration of the dataset.

tains the time-curve of each parameter from all the three resolutions plotted together. Based on user's selection, all the parameters from the selected days and resolutions are highlighted in the line chart. The MDS view and the line chart view are also connected to the spatial views (discussed in Section 4.2.1) for detailed spatio-temporal analysis of the sensitivity values.

5 Accuracy Analysis Across Resolutions

After analyzing the sensitivity of the input parameters of the simulation model, the other important part is to understand the accuracy of the simulated outputs. Since the ensemble datasets are generated at different resolutions, understanding the effect of the resolution is of great interest to the domain experts. Although it is generally expected that the higher resolution simulation output will be able to predict the observed phenomenon much better than the lowest resolution, it is not necessarily true in the real ensemble simulations. With the increase in resolution, the model uncertainty can also increase for a simulation like the precipitation of WRF model. Domain experts need to understand these variations and intend to get information for future adaptive grid implementations. In this section, we discuss our proposed solution to handle this scenario for helping the domain expert with an effective exploration tool.

5.1 Bayesian Model For Accuracy Comparison

Given an ensemble dataset and its observed ground truth, Bayesian Model Averaging (BMA) [10] was previously employed to understand the accuracy of the multiple ensemble runs. Since the dataset we are dealing with is multi-resolution, we formulate a different strategy for accuracy analysis. We use a Bayesian approach across the three resolutions to find the estimated best predictor. Given a time t and location $\vec{v}$, our simulation model generates a collection of precipitation estimates for every resolution. Let d_i represent the distribution of the rainfall values generated by the ith resolution at a given location at a given time and R is the random variable for the estimated rainfall values. Let D be the random variable that represents the user's choice of which distribution d_i is selected from the k distributions originating from the k resolutions. Thus, $P(D = d_i)$ represents the probability of choosing the ith distribution from the k distributions. Now, at a given location with the observed rainfall amount r, we want to identify the best local predictor as the ith resolution when the conditional probability $P(D = d_i|R = r)$ is maximized:

$$\underset{i \in k}{\operatorname{argmax}}\ P(D = d_i|R = r). \tag{8}$$

From Bayesian formulation, it can be readily observed that:

$$P(D = d_i|R = r) \propto P(R = r|D = d_i)P(D = d_i). \tag{9}$$

In this equation, $P(D = d_i)$ is the prior unconditional probability of choosing ith resolution as the best predictor. In the absence of the prior information, $P(D = d_i)$ can be set to equally probable for all the resolutions, i.e., $P(D = d_i) = \frac{1}{k}, \forall i \in k$. Depending on domain expert's choice, this probability can be set to be inversely proportional to the computation and storage cost of the ith resolution data set, i.e., $P(D = d_i) \propto Cost(i) \Rightarrow P(D = d_i) = \varepsilon * Cost(i)$. Domain experts can tune this prior probability for understanding the cost vs accuracy benefits of multiple resolutions.

5.2 Kernel Density Estimation For Ensembles

The conditional probability $P(R = r|D = d_i)$ of Equation 9 intuitively captures the likelihood of distribution d_i generating the sample r. Although Gaussian assumption is popular among the weather analysts, in this work we take up a non-parametric approach and use kernel density estimation (KDE) to approximate the distribution of the ensemble outputs at every spatial location. Although slower, this KDE approach is beneficial over the Gaussian approach since: 1) it is parameter free unlike Gaussian estimation, 2) Gaussian distributions are known to produce non-negligible likelihood of getting negative values whereas our precipitation data is strictly positive. For KDE, the popular choices of kernel types include Gaussian, triangular, rectangular, and the Epanechnikov kernel. Different choices of the kernel types produce different estimated density functions although the variation due to kernel types is considered less significant compared to the choice of kernel bandwidth [33, 38]. When the Gaussian kernel is used, Silverman's rule of thumb for bandwidth selection produces a good quality density estimation. Following this method, bandwidth h is given as $h = (\frac{4}{d+2})^{1/(d+4)} \sigma n^{-1/(d+4)} = 1.05\sigma n^{-1/5}$ where σ is the standard deviation of the ensembles at a given location, n is the number of ensemble runs and d is the data dimensionality ($d = 1$ in our case). The kernel was chosen to be the Gaussian kernel in this work with this method of bandwidth selection.

Using the KDE approach at every spatial location, three distributions can be estimated from the ensembles coming from the three resolutions. From these density functions, the local best predictor is identified by approximating the conditional probability $P(R = r|D = d_i)$ as the likelihood of r from the corresponding density functions. Thus, $P(D = d_i|R = r)$ can be approximated as Equation 9 and comparing the values from three resolutions, the best local predictor can be identified. Since the densities estimated from KDE are used for comparison, the same bandwidth needs to be used for a spatial location across all the three resolutions while computing the d_i's. After identifying the local best predictors for each time step, our goal is to explore the prediction performance of the three resolutions across time domain for different spatial locations. To achieve this, we find the three tuple $\{e_h, e_m, e_l\}$ that represents the temporal prediction performance of the three resolutions at a given location. Thus, e_h denotes the count of how many times the highest resolution was chosen as the best predictor at the given location over all the days; similarly, e_m and e_l denote the counts for middle and lowest resolutions respectively. Finally, we turn the counts into corresponding probability values $\{p_h, p_m, p_l\}$ by dividing the counts by the total number of days. Thus, p_h now represents the probability of selecting the highest resolution as the best predictor over all the days at the given location. After this conversion from the ensemble data to probability values, the resulting dataset can

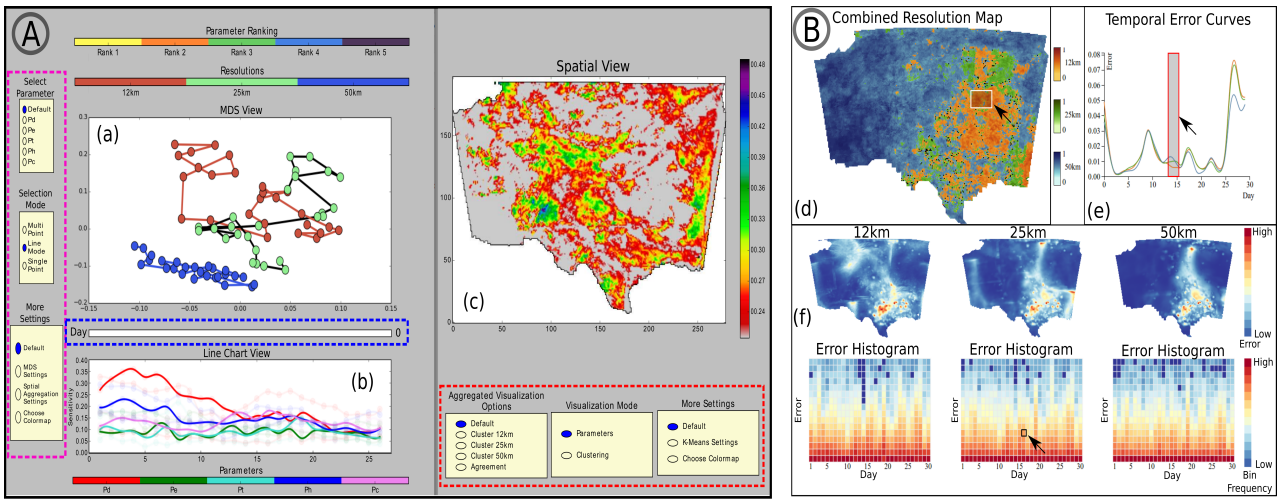


Fig. 8: Visualization of our system: (A) sensitivity exploration mode, (B) accuracy analysis mode.

be used for visualization.

5.3 Visualization of Multi-resolution Efficiency

To use the previously generated probability data, a visualization is needed that easily and effectively transfers the information to the domain expert. Since there is a tuple $\vec{p} = \{p_h, p_m, p_l\}$ at every location of a two-dimensional field, one option is to directly map the probability values to r, g, b values for visualization. But this produces an ineffective visualization as shown Figure 7a. Instead, we first assign three disjoint color-maps to represent the three resolutions and depending on which resolution has the highest probability at a given location, the final color is selected from the associated color-map. In essence, we first classify each location based on the probability values and then use the largest probability value to modulate the color. Since different colors have different brightness to the human eye and their contrasts are perceived differently, special care is needed in choosing the color-maps. In this work, we collaborated with an artist who has more than twenty years of experience in the art and visualization domain to select the three color-maps for the three resolutions. These three color-maps are presented in Figure 7b and using this custom color coding, final output is generated as shown in Figure 7c. From this aggregated image, domain experts get an idea of where the different resolutions are performing better than the others. Comparing this image with the mean observed data as shown in Figure 1c, it is readily observed that the highest resolution ensemble performs quite well in the regions of higher rainfall. Similarly, the lowest resolution predicts the dry regions better than the other two resolutions.

6 System Integration and User Interaction

In this section, we integrate all the system components together to help the users explore the time-varying multi-resolution ensemble datasets. The full system view and user interactions are provided in the accompanying video of this paper. Here we discuss some key concepts regarding our system. A snapshot of the system for sensitivity analysis is shown in Figure 8A where we have three connected visualization windows, Window (a),(b),(c). In Window (a), the MDS plot of the spatially aggregated data points (representing different days) are displayed as discussed in Section 4.2.2. In the "Default" view of the "Select-Parameter" option of the selection panel (marked by violet outline), the MDS points are color-mapped to represent the different resolutions to provide a high-level view of how the three resolutions are related over the temporal domain. From the selection panel of "Select-Parameter", users can select an input parameter to explore the rank and sensitivity of it across all the three resolutions similar to Figure 6. In this mode, the point sizes are scaled according to the sensitivity values of the selected parameter and the color of points reflect the rank (brighter colors represent higher ranks according to a custom color-map shown at the top of window (A)). Users are provided with a "Day" slider (marked by the blue outline) to interactively highlight different points of the MDS plot across all the resolutions for sensitivity-based similarity comparison. From the left selection panel, users can operate in "Multi-Point" mode to select any number of days by drawing a free-hand curve around the points. The line chart view of Window (b) will then be updated to show the sensitivity values of all the five parameters for the user-selected days. When in "Line-Mode", users can select one of the three resolutions from the MDS view to observe the temporal sensitivity trends of the five parameters as time-curves in the line chart window. In Figure 8, 25km resolution is selected and the sensitivity trends are shown in Window (b). From the "Single-Point" mode, users can select a day from Window (a) for spatial exploration of the selected parameter in Window (c). For spatial view, Window (c) has two modes of visualization: 1) parameter sensitivity distribution across spatial locations, 2) clustering information for different spatial locations. When the "Parameters" view is selected, it shows the spatial distribution of δ values of the selected parameter for the user selected day. In this mode, the users can also interactively set the color-map of the visualization. When the "Clustering" view is selected, it shows how the clustering based on the input parameter sensitivities (as discussed in Section 4.2.1) vary across spatial regions for the user selected day. To provide a time-aggregated view of the clustering, the user can also select to view the clustering information of the temporally averaged data for the three resolutions similar to Figure 4. In this interface, users can alter the system settings (shown in red and violet outlines) and continue to explore in detail how the sensitivity values of the input parameters change across spatial locations, temporal domain and three resolutions.

In the accuracy exploration mode, as shown in Figure 8B, we first present the combined resolution view of the dataset in Window (d) using our custom color-map (method discussed in Section 5). Users can interactively use brushing-and-linking to explore the regions of Window (d) for further analysis. Since our combined map view only presents which resolution is performing better at which spatial locations, we add the error analysis to augment the visualization. In this case, we are using the popular root-mean-square-error (RMSE) measure for each point in the spatio-temporal domain. Further, the RMSE errors are spatially aggregated for each day to generate "temporal error curves" for each of the resolutions in Window (e). These simultaneous curves show a comparative view of how much error each resolution produces in predicting the observed rainfall over the temporal domain. When a rectangular region from Window (d) is selected (marked with a white box), the error curves of Window (e) are updated to show the temporal trend of error for the user-selected region. Users can also select an interval of days from Window (e) (marked by red box) and further explore how the error is distributed over the spatial domain for the three resolutions for those user-selected days in the connected

spatial-error views provided at the top panel of the Window (f). For the users to quickly identify the regions of high error from the three resolutions, we provide an error histogram panel below the spatial-error view of the individual resolutions of Window (f). In these error histograms, the RMSE errors of each day are separated into user defined k-bins. The x-axis represents the different days and the y-axis shows the increasing error values. The height of each bin is mapped to color according to a user-defined color-map. This settings helps the users identify the high frequency and/or high error bins easily. Users can select an error bin (marked with the black box) from one resolution, and the spatial locations contributing to that bin are highlighted in the connected spatial-error map view. The error generated in the other two resolutions for these spatial locations are also highlighted in the corresponding spatial views. Using these interactions and methodology, our designed system can provide the users a comprehensive view of the accuracy and error in the simulation generated output across multiple resolutions.

7 Domain Expert Feedback

We collaborated with an expert of the field with multiple years of experience in this domain. Our proposed system was developed after a series of thorough and detailed discussions with the expert. The domain specific requirements were communicated in multiple phases, thereby refining the tools and methods that are to be used in the final software. The domain expert was involved at every stage of the pipeline to ensure that the goals of the experiments were achieved. Expert feedback was obtained during the several meetings that we organized with the domain expert throughout the development phase and after the tool was completed. The final comments of the expert are listed here that point to the benefits of the our system and future improvements.

7.1 Feedback On The System And Methodology

The feedback we received from the expert was very encouraging. The expert commented that our designed tool enabled him to explore the sensitivity information at different levels of data aggregation. This is useful to him specially because a lot of averaging is done by the experts on the weather simulation output to predict and understand the climate situations over different time-domain and spatial domain. Our proposed tool provided the methodology to walk-through the different data aggregations with simple user inputs and interactions. The expert did not have an existing visualization tool that explores these aspects of the input parameters. He also mentioned that the accuracy analysis tool effectively depicts the performance of the different resolutions with simple user interactions. Our system generated images clearly establish that the highest resolution simulation does not always perform better than the lower resolutions, which is counter-intuitive but important knowledge for the domain expert. Since our tool effectively highlights the spatial and temporal behavioral patterns of different regions, the expert was able to quickly locate the days and regions of interest from the tool. Since this software was cross-domain, we kept the interactions easy and intuitive with intuitive color-maps. The ease of use was another aspect that the domain expert agreed on. Overall, the expert expressed that the system met the goals for the project and all the components used in the system were needed for the system to perform well.

7.2 Domain Specific Discoveries

Our system helped the expert in making several domain specific discoveries. The MDS view revealed that the lowest resolution has very different sensitivity values on the different days compared to the other two higher resolutions. The MDS view connected with the line charts also reveals that the input parameter *Pd* which is the coefficient related to downdraft mass flux rate, is the most important input parameter in determining the output precipitation for most of the days of high and middle resolutions. For the lowest resolution, all the input parameters are very similar with low sensitivity values. This was a new discovery for the dataset which was revealed due to the detailed sensitivity analysis. The clustering based spatial view revealed that for the overall days, the different resolutions agree with the sensitivity of the output precipitation to the input parameters for the regions that are either very wet or very dry. Also, from individual clustering visualizations, it was revealed that the lowest resolution was again very different from the two higher resolutions in the spatial clustering information as well. Similarly, in the accuracy exploration, the fact that different regions are well predicted by different resolutions was very crucial information for the domain expert. Although it was expected that the higher resolution may not always provide better prediction compared to the lower resolutions, the performance of the lowest resolution in the less wet and dry regions compared to the other two resolutions, was surprising to the expert. The domain expert later concluded that the reason behind this observation is the model uncertainty contained in the WRF climate model. Due to this, when increasing the model resolution, the model may become more unstable and cause larger model uncertainty. This may result in higher inaccuracy for the higher resolution in certain areas compared to the lower one. Also, our expert expressed that the knowledge of high resolution as a better predictor for the more wet regions will guide the expert towards a better adaptive grid refinement in the future.

7.3 Future Improvements And Updates

The domain expert also provided us with suggestions for improving our system. Instead of incorporating only the δ measure for sensitivity analysis, the expert wanted us to include more sensitivity measures to understand how they rank the different parameters. Understanding the consensus among the parameters can help us predict the parameter sensitivity more robustly. In the accuracy analysis, the domain expert wanted us to develop an AMR grid based interactive cost optimization strategy that can be user controlled. Here the cost refers to the computational cost of the model. The domain expert also wanted us to extend our visual exploration system to *in situ* workflow and also to handle three dimensional multi-resolution temporal ensembles. Finally, we plan to involve multiple domain experts for further improvements and feedback on our system. We take these suggestions from the domain expert as our future work.

8 Conclusion

In this paper, we presented a complete system to explore multi-resolution temporal weather ensembles. Our system is intended for exploring the sensitivity of the input parameters and accuracy of the output precipitation across resolutions. In this work, we used a global sensitivity measure named δ to compute the sensitivity of the different input parameters over spatial regions and over temporal domain for the three given resolutions. To explore how sensitivity of the parameters change over spatial regions, a location-based clustering is applied for different days and the clustering-agreement among the three resolutions is computed. For analyzing temporal patterns, spatially aggregated dataset is used for sensitivity computation given all the parameters and an MDS plot linked with line charts is presented for user interaction. For accuracy exploration, a Bayes rule based likelihood method is used to compute the local best predictor among the three resolutions for every spatial location from the ensembles. To create local probability density functions for this likelihood comparison, KDE approach is used. The local best predictors are then aggregated in the temporal domain to find which location is better predicted by which resolution over the total number of days and it is color-mapped for further user interaction. Users can now select regions from this image and analyze the temporal and spatial error pattern in a linked window. We carried out the experiments in close collaboration with an expert whose feedback establishes the efficacy of our proposed system.

Acknowledgments

The authors wish to thank Francesca Samsel for the color-maps and the anonymous reviewers for their insightful comments and suggestions. This work was supported in part by NSF grants IIS-1250752, IIS-1065025, and US Department of Energy grants DE-SC0007444, DE-DC0012495, program manager Lucy Nowell.

REFERENCES

[1] K. Bensema, L. Gosink, H. Obermaier, and K. Joy. Modality-driven classification and visualization of ensemble variance. *IEEE Transactions on Visualization and Computer Graphics*, PP(99):1–1, 2015.

[2] A. Bock, A. Pembroke, M. Mays, L. Rastaetter, T. Ropinski, and A. Ynnerman. Visual verification of space weather ensemble simulations. In *2015 IEEE Scientific Visualization Conference*, pages 17–24, Oct 2015.

[3] E. Borgonovo. A new uncertainty importance measure. *Reliability Engineering and System Safety*, 92(6):771–784, 2007.

[4] N. E. Bowler, A. Arribas, K. R. Mylne, K. B. Robertson, and S. E. Beare. The MOGREPS short-range ensemble prediction system. *Q.J.R. Meteorol. Soc.*, 134(632):703–722, apr 2008.

[5] D. G. Cacuci. Sensitivity theory for nonlinear systems. i. nonlinear functional analysis approach. *Journal of Mathematical Physics*, 22(12), 1981.

[6] W. Castaings, D. Dartus, F.-X. Le Dimet, and G.-M. Saulnier. Sensitivity analysis and parameter estimation for distributed hydrological modeling: potential of variational methods. *Hydrology and Earth System Sciences*, 13(4):503–517, 2009.

[7] T. F. Cox and M. A. A. Cox. *Multidimensional Scaling*. Chapman & Hall, London, 2001.

[8] I. Demir, C. Dick, and R. Westermann. Multi-charts for comparative 3d ensemble visualization. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2694–2703, Dec 2014.

[9] F. Ferstl, K. Brger, and R. Westermann. Streamline variability plots for characterizing the uncertainty in vector field ensembles. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):767–776, Jan 2016.

[10] L. Gosink, K. Bensema, T. Pulsipher, H. Obermaier, M. Henry, H. Childs, and K. I. Joy. Characterizing and visualizing predictive uncertainty in numerical ensembles through bayesian model averaging. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2703–2712, Dec 2013.

[11] H. Guo, X. Yuan, J. Huang, and X. Zhu. Coupled ensemble flow line advection and analysis. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2733–2742, Dec 2013.

[12] Z. Guo, M. Wang, Y. Qian, V. E. Larson, S. Ghan, M. Ovchinnikov, P. A. Bogenschutz, C. Zhao, G. Lin, and T. Zhou. A sensitivity analysis of cloud properties to clubb parameters in the single-column community atmosphere model (scam5). *Journal of Advances in Modeling Earth Systems*, 6(3):829–858, 2014.

[13] J. Hacker, S. Ha, C. Snyder, J. Berner, F. Eckel, E. Kuchera, M. Pocernich, S. Rugg, J. Schramm, and X. Wang. The u.s. air force weather agencys mesoscale ensemble: scientific description and performance results. *Tellus A*, 63(3), 2011.

[14] L. Hao, C. G. Healey, and S. A. Bass. Effective visualization of temporal ensembles. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):787–796, Jan 2016.

[15] T. Hollt, A. Magdy, P. Zhan, G. Chen, G. Gopalakrishnan, I. Hoteit, C. D. Hansen, and M. Hadwiger. Ovis: A framework for visual analysisof ocean forecast ensembles. *IEEE Transactions on Visualization and Computer Graphics*, 20(8):1114–1126, Aug 2014.

[16] Z. Hou, M. Huang, L. Leung, G. Lin, and D. Ricciuto. Sensitivity of surface flux simulations to hydrologic parameters based on an uncertainty quantification framework applied to the community land model. *Journal of Geophysical Research. D. (Atmospheres)*, 117(D15):D15108, 2012.

[17] T. Hllt, A. Magdy, G. Chen, G. Gopalakrishnan, I. Hoteit, C. D. Hansen, and M. Hadwiger. Visual analysis of uncertainties in ocean forecasts for planning and operation of off-shore structures. In *Visualization Symposium (PacificVis), 2013 IEEE Pacific*, pages 185–192, Feb 2013.

[18] R. Ibbitt and T. ODonnell. Sensitivity theory for nonlinear systems. i. nonlinear functional analysis approach. *IAHS Publication*, 101:462–475, 1971.

[19] B. Iooss and P. Lemaître. *Uncertainty Management in Simulation-Optimization of Complex Systems: Algorithms and Applications*, chapter A Review on Global Sensitivity Analysis Methods, pages 101–122. Springer US, 2015.

[20] M. Jarema, I. Demir, J. Kehrer, and R. Westermann. Comparative visual analysis of vector field ensembles. In *Visual Analytics Science and Technology (VAST), 2015 IEEE Conference on*, pages 81–88, Oct 2015.

[21] C. Johnson and A. Sanderson. A next step: Visualizing errors and uncertainty. *IEEE Comp. Graphics and Applications*, 23(5):6–10, Sept 2003.

[22] P. R. Johnston and D. H. Pilgrim. Parameter optimization for watershed models. *Water Resources Research*, 12(3):477–486, 1976.

[23] F. Liang, Y. Cheng, and G. Lin. Simulated stochastic approximation annealing for global optimization with a square-root cooling schedule. *Journal of the American Statistical Association*, 109(506):847–863, 2014.

[24] W. J. C. M. D. McKay, R. J. Beckman. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245, 1979.

[25] M. Mirzargar, R. T. Whitaker, and R. M. Kirby. Curve boxplot: Generalization of boxplot for ensembles of curves. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2654–2663, Dec 2014.

[26] M. D. Morris. Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33(2):161–174, 1991.

[27] T. A. Pang, M. C. Wittenbrink, and K. S. Lodha. Approaches to uncertainty visualization. *The Visual Computer*, 13(8):370–390, 1996.

[28] T. Pfaffelmoser, M. Reitinger, and R. Westermann. Visualizing the positional and geometrical variability of isosurfaces in uncertain scalar fields. In *Computer Graphics Forum*, volume 30, pages 951–960, 2011.

[29] K. Pothkow and H. C. Hege. Positional uncertainty of isocontours: Condition analysis and probabilistic measures. *IEEE Transactions on Visualization and Computer Graphics*, 17(10):1393–1406, Oct 2011.

[30] K. Pöthkow, B. Weber, and H.-C. Hege. Probabilistic marching cubes. *Computer Graphics Forum*, 30(3):931–940, 2011.

[31] K. Potter, P. Rosen, and C. R. Johnson. *Uncertainty Quantification in Scientific Computing: 10th IFIP WG 2.5 Working Conference*, chapter From Quantification to Visualization: A Taxonomy of Uncertainty Visualization Approaches, pages 226–249. Springer Berlin Heidelberg, 2012.

[32] K. Potter, A. Wilson, P.-T. Bremer, D. Williams, C. Doutriaux, V. Pascucci, and C. Johhson. Visualization of uncertainty and ensemble data: Exploration of climate modeling and weather forecast data with integrated visus-cdat systems. *Journal of Physics: Conference Series*, 180(1):012089, 2009.

[33] K. Pthkow and H.-C. Hege. Nonparametric models for uncertainty visualization. *Computer Graphics Forum*, 32(3pt2):131–140, 2013.

[34] A. Saltelli. Sensitivity analysis for importance assessment. *Risk Analysis*, 22(3):579–590, 2002.

[35] A. Saltelli, S. Tarantola, and K. P.-S. Chan. A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics*, 41(1):39–56, 1999.

[36] J. Sanyal, S. Zhang, J. Dyer, A. Mercer, P. Amburn, and R. Moorhead. Noodles: A tool for visualization of numerical weather model ensemble uncertainty. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1421–1430, Nov 2010.

[37] S. Schlegel, N. Korn, and G. Scheuermann. On the interpolation of data with normally distributed uncertainty for visualization. *IEEE Trans. on Visualization and Computer Graphics*, 18(12):2305–2314, Dec 2012.

[38] B. W. Silverman, editor. *Density estimation for statistics and data analysis*. Chapman & Hall/CRC, 1992.

[39] I. Sobol and S. Kucherenko. Derivative based global sensitivity measures and their link with global sensitivity indices. *Mathematics and Computers in Simulation*, 79(10):3009 – 3017, 2009.

[40] I. Sobol. Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. *Mathematics and Computers in Simulation*, 55(13):271 – 280, 2001.

[41] M. van Lier-Walqui, T. Vukicevic, and D. J. Posselt. Quantification of cloud microphysical parameterization uncertainty using radar reflectivity. *Mon Weather Rev*, 140, 2012.

[42] G. Warren, S and H. Schneider S. Seasonal simulation as a test for uncertainties in the parameterizations of a budyko-sellers zonal climate model. *J Atmos Sci*, 36:1377–1391, 1979.

[43] R. T. Whitaker, M. Mirzargar, and R. M. Kirby. Contour boxplots: A method for characterizing uncertainty in feature sets from simulation ensembles. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2713–2722, Dec 2013.

[44] H. Yan, H. Qian, G. Lin, L. Leung, B. Yang, and Q. Fu. Parametric sensitivity and calibration for the kain-fritsch convective parameterization scheme in the wrf model. *Climate Research*, 59:135–147, 2014.

[45] B. Yang, Y. Qian, G. Lin, R. Leung, and Y. Zhang. Some issues in uncertainty quantification and parameter tuning: a case study of convective parameterization scheme in the wrf regional climate model. *Atmospheric Chemistry and Physics*, 12(5):2409–2427, 2012.

[46] C. Zhao, X. Liu, Y. Qian, J. Yoon, Z. Hou, G. Lin, S. McFarlane, H. Wang, B. Yang, P.-L. Ma, H. Yan, and J. Bao. A sensitivity study of radiative fluxes at the top of atmosphere to cloud-microphysics and aerosol parameters in the community atmosphere model cam5. *Atmospheric Chemistry and Physics*, 13(21):10969–10987, 2013.