1.9 APPLICATION OF EVIDENCE THEORY TO QUANTIFY UNCERTAINTY IN FORECAST OF HURRICANE PATH

Svetlana V. Poroseva, Julie Letschert** and M. Yousuff Hussaini
School of Computational Science, Florida State University, FL 32306-4120, USA

1. INTRODUCTION

Results of any computations are of practical use only if information about their accuracy is also available. This is especially true in forecasting hurricane paths, where the prediction accuracy is of vital importance. Yet, most current forecasts lack such information

One way to assess the accuracy of a forecast is to identify and describe all sources of uncertainty (see Oberkampf et al. 2004, and the references therein.) The attractiveness of this idea is clear. If one could identify and describe all uncertainty sources, contributions from at least some of them could be eliminated or reduced, thereby making simulations more credible. Uncertainties and errors in computational results on hurricane forecasts originate from various sources -- failure of a climate model to correctly describe the atmospheric physics and the interaction between the atmosphere and the ocean; stochastic nature of model parameters; errors associated with the discretization and algorithmic approximations, to mention just a few. It could be an impossible task to identify all uncertainty sources though. Even if some sources of uncertainty are identified, there still remains the problem of describing their contributions, as these sources might affect one another in complicated and generally unknown ways. For instance, decreasing the contribution from one source can increase uncertainty from other sources, resulting in increased total uncertainty. In this sense, one has to be aware that incorporating additional model equations in a climate model for the sake of a better description of physics may not always result in more accurate forecasts. Uncertainty in computational results caused by new equations and parameters can exceed benefits of a better physical model. Evaluation of the effectiveness of modifications introduced in a climate model and quantitative comparison of the performance of different models requires a measure for quantifying the total uncertainty in hurricane forecasts, that is, the sum of contributions from all uncertainty sources. To find such a measure is one of the objectives of our research.

The total uncertainty in flow simulations has features of both aleatory and epistemic uncertainties currently recognized. Aleatory uncertainty is due to stochastic influences (e.g., random noise) and cannot be reduced. Epistemic uncertainty is subjective and originates from incomplete knowledge at any stage of modeling or simulation. Increasing one's knowledge reduces epistemic uncertainty. As both aleatory and epistemic uncertainties are intricately interwoven in a hurricane forecast, one needs a statistical theory that can handle them together to quantify their impact on forecasts.

There are several mathematical theories that describe uncertainty and provide its measures -- probability theory, possibility theory, and evidence theory. Probability theory, for instance, is better suited to describe aleatory uncertainty. Possibility theory (Dubois&Prade 1988) was developed mainly to describe epistemic uncertainty. An extensive literature exists (see e.g., Klir& Wierman 1998), where various uncertainty theories are compared, their relations are established, and their advantages and limitations are discussed. Evidence theory (Shafer 1976a) is among the well-established theories that can handle both

* *Corresponding author address:* myh@csit.fsu.edu

**Undergraduate intern from National Institute of Applied Sciences (I.N.S.A.), Lyon, France.

types of uncertainty and does not require their separation. In fact, probability and possibility theories are branches of evidence theory. The theory works with limited information and new data can be incorporated as it becomes available. These features make evidence theory attractive for application to hurricane forecasts, and the present study investigates the potential of using evidence theory to provide a quantitative assessment of forecast accuracy. Unfortunately, there are very few practical applications of evidence theory and they differ considerably from the one addressed here. Previously, we developed an approach based on evidence theory to quantify uncertainty in turbulence computations (Poroseva et al. 2005). Results of testing the approach on a turbulent flow encouraged us to apply a similar approach to hurricane path forecasts with appropriate extension and modification.

Evidence theory provides the necessary tools not only to quantify the forecast uncertainty, but also to fuse the results of different forecasts. To develop a reliable procedure for combining different forecasts based on evidence theory tools is another objective of our study. The idea of improving the overall credibility of hurricane path predictions by combining results of several forecasts is not new. The multimodel superensemble technique (Williford et al. 2003) is an example of the successful implementation of the idea. However, multimodel forecasts, like the single model forecasts, do not provide information on the forecast accuracy. The present approach provides the quantitative assessment of the forecast accuracy and differs completely from other multimodel techniques in its mathematical foundation.

2. MATHEMATICAL BACKGROUND

In this paper, we follow the axiomatic approach of evidence theory given by Shafer (1976a). A comprehensive exposition of the foundations of evidence theory may also be found in Sentz & Ferson (2002), Shafer (1976b, 1987, 1990), and Yager et al. (1994). In the interest of space, a brief description of the basic concepts of the theory is provided here for the sake of completeness.

Evidence theory provides two basic tools for quantifying uncertainty in simulations and improving predictions: i) a tool for representing the degree of belief (confidence) that may be attributed to a given proposition on the basis of given evidence, and ii) a tool for combining evidence from different sources (Dempster's rule). Let X denote a quantity and $\mathbf{X}$ the finite set of its possible values. Then, propositions can be of the form "the true value of X is in A," where A is a subset of $\mathbf{X}$. Whenever A is interpreted as a proposition, its complement $\overline{A}$ (the set of all elements of $\mathbf{X}$ not in A) must be interpreted as the proposition's negation. The set of all subsets of $\mathbf{X}$, the power set, includes the empty set $\varnothing$ (corresponding to a necessarily false proposition, since the true value cannot lie in $\varnothing$) and the entire set $\mathbf{X}$ (corresponding to a necessarily true proposition, since the true value is assumed to be in $\mathbf{X}$).

In evidence theory, the impact of evidence on our belief in different propositions is described by three related functions -- the basic probability assignment function (m), the belief function (Bel), and the plausibility function (Pl). The basic probability assignment function assigns a number $m(A)$ to each subset A of $\mathbf{X}$ such that $m(\varnothing)=0$ for the empty set $\varnothing$, and the sum of basic probability assignments (BPAs) for all subsets A of $\mathbf{X}$ is equal to unity:

$$\sum_{A\subseteq \mathbf{X}} m(A)=1. \qquad (1)$$

The quantity $m(A)$ is the measure of the belief that is committed exactly to A but not to any particular subset of A. The belief in A is based on available evidence that supports exactly A. As $m(A)$ is a measure of the belief committed exactly to A, it does not represent the total belief committed to A. In evidence theory, a measure of the total belief (degree of belief) in A is defined as

$$Bel(A)=\sum_{B\subset A} m(B), \qquad (2)$$

reflecting the fact that the evidential support committed to one proposition is committed to any subset containing it. A subset *A* of **X** is called a focal element of a belief function *Bel* over **X** if $m(A) > 0$. The union of all focal elements of a belief function is called its core. The plausibility measure is related to the basic probability assignment *m*:

$$Pl(A) = \sum_{B \cap A \neq \varnothing} m(B) . \qquad (3)$$

Belief and plausibility measures are related by the equation $Pl(A) = 1 - Bel(\overline{A})$. Some properties of these measures are

$$Bel(\varnothing) = Pl(\varnothing) = 0 ,$$
$$Bel(\mathbf{X}) = Pl(\mathbf{X}) = 1 ,$$
$$Bel(A) \leq Pl(A) ;$$
if $B \subseteq A$, then $Bel(B) \leq Bel(A)$ and $Pl(B) \leq Pl(A)$,
$$Bel(A) + Bel(\overline{A}) \leq 1 \text{ and } Pl(A) + Pl(\overline{A}) \geq 1 .$$

The last two expressions show that the two measures are nonadditive, that is, the sum of belief measures and the sum of plausibility measures are not required to be equal to unity. It is a consequence of uncertainty in available evidence. When evidence supports with certainty mutually exclusive propositions, the two measures coincide and the additivity rule is recovered.

Notice that the way one defines subsets *A* of **X** and links actual evidence to their basic assignments $m(A)$ depends on the problem being considered, one's current limited knowledge, and available evidence. Additional information can change the set of propositions and how evidence determines our degree of belief $Bel(A)$ in these propositions.

Dempster's rule is a technique for combining evidence from different sources to improve predictions. Mathematically, application of Dempster's rule to two or more belief functions over the same set **X** yields a new belief function called their orthogonal sum. In the simplest case of two belief functions Bel_1 and Bel_2 with basic probability assignment functions m_1 and m_2, Dempster's rule provides the orthogonal sum

$$m(C) = \frac{\sum_{\substack{i,j \\ A_i \cap B_j = C}} m_1(A_i) m_2(B_j)}{1 - \sum_{\substack{i,j \\ A_i \cap B_j = \varnothing}} m_1(A_i) m_2(B_j)} , \qquad (4)$$

where $C = A_i \cap B_j$; $A_1, ..., A_k$ and $B_1, ..., B_k$ are focal elements of Bel_1 and Bel_2, respectively. The core of the belief function given by *m* is equal to the intersection of the cores of Bel_1 and Bel_2. The belief function $Bel(C)$ resulting from the combination can then be obtained from *m* using Equation (2).

For valid use of Dempster's rule, belief functions Bel_1 and Bel_2 must satisfy some conditions (Shafer 1976a): they should not strictly contradict each other and they should be based on independent sources of evidence. We discuss how to satisfy these requirements for the specific problem being considered in the paper in the following sections.

3. EVALUATION OF MODEL ACCURACY

In this section we describe a new technique based on the basic concepts of evidence theory to evaluate the accuracy of a climate model using available observational databases for hurricane paths. Specifically, uncertainty is quantified in terms of a basic probability assignment function (*m*-function) for intervals in which the deviation of computed results from observational data falls.

The database for hurricanes from the years 1998-2001 in the Pacific Ocean is used to quantify uncertainty in forecasts. The database consists of observational data for the hurricane position – latitude and longitude – at different instants and of the predictions by several global models of the positions of the same hurricanes. In the current study, we are using the results produced by models from two operational centers -- the U.S. Navy Operational Global Atmospheric Prediction System (NOGAPS)

and the European Centre for Medium-Range Weather Forecasts (ECMRWF).

The accuracy of the model prediction is evaluated by comparing the observational data for the latitude/longitude with the model predictions at a given instant. The difference between observational and model data,

$$Dev = L^o - L^m , \qquad (5)$$

is called the deviation (*Dev*), where L^m denotes the model prediction of the latitude or longitude; and L^o is the corresponding observational value. All quantities are measured in degrees.

Since a model can produce several independent forecasts for the same hurricane, we have to define instants at which observational and model data are compared and deviation values are obtained. We denote as t_i^o an instant at which observational data for the hurricane position are available, where $i = 1,...,s$, and *s* is the total number of such instants for a given hurricane. The time step (the interval between two instants) is 12 hours (12*h)* in the current study.

A new forecast can be produced with different frequency; for instance, each 6 hours (6*h*), 12*h* or each day (24*h*). In this study, we use data of forecasts produced every twelve hours. That is, the second forecast (*f*2) started 12*h* later than a hurricane itself; the time difference between third (*f*3) and second (*f*2) forecasts is 12*h*, etc. Forecast instants corresponding to observational data are denoted as t_j^{fk} , where $j = 1,...,r$, *r* is the amount of time steps in a given forecast; *fk* is an ordinal number of a forecast: *fk* = *f*1, *f*2,…) Since at this stage, we compare forecasts with the observational data of hurricanes that have already occurred (that is, the duration of a hurricane is known), the duration of each consequent forecast decreases. For the first forecast of a hurricane, the number of time steps is the same as in the hurricane: $t^o_{1,...,s} = t^{f1}_{1,...,s} + 0h$, and $r = s$. The *k*-th forecast is shifted by $12 \cdot (k-1)$ hours from the start of the hurricane: $t^o_{k,...,s} = t^{fk}_{1,...,s-k} + 12 \cdot (k-1)h$, therefore $r = s - k$ and the first instant of the forecast corresponds to the *k*-th instant of the observational data.

Thus, for the same hurricane, we can obtain several deviation values (*Dev*-values) that characterize the deviation of the model data from observational data at a given instant of a forecast. For instance, for the initial moment of a forecast ($t_1 = t_1^{f1} = t_1^{f2} = ... = 0h$), there are as many *Dev*-values as there are forecasts for this hurricane. These *Dev*-values are defined as

$$Dev(t_1^{fk}) = L^o(t_k^o) - L^m(t_1^{fk}) .$$

At the next forecast instant, $t_2 = t_1 + 12h = 12h$,

$$Dev(t_2^{fk}) = L^o(t_{k+1}^o) - L^m(t_2^{fk}) ,$$

and so on. Since, the duration of forecasts decreases with increasing *k*, the number of *Dev*-values corresponding to different forecast instants also decreases. We denote as $Dev(t_1)$ the set of all available $Dev(t_1^{fk})$ -values, $Dev(t_2)$ the set of $Dev(t_2^{fk})$ -values etc.; there are *s* such sets for a given hurricane of *s* time steps. Each set carries the information on model accuracy at a given forecast instant. We can further increase the population (the number *N* of *Dev*-values) of each deviation set (*Dev*-set) by adding the *Dev*-values related to all available hurricanes of a given year at a given forecast instant. Table 1 shows, as an example, the populations of the *Dev*-sets at different instants corresponding to the forecasts produced by the ECMRWF model in the years 1998-99.

t_i , *h*	*N*, 1998	*N*, 1999
0	225	302
12	210	270
24	180	226
36	161	192
48	134	161
60	110	133
72	93	112
84	78	95
96	64	77
108	58	64
120	51	52

TABLE 1. The number of *Dev*-values in *Dev*-sets at different forecast instants obtained for the ECMRWF model results in the years 1998-99.

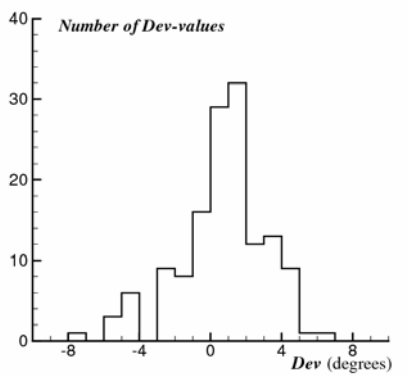


FIGURE 1. Example of *Dev*-distribution for the latitude at $t_6 = 60h$ obtained with $\Delta Dev = 1^o$. (ECMRWF model, 2001).

Dev-sets can be used to characterize the accuracy of a model at each instant. The procedure will be discussed below in relation to a single instant and a single quantity, the latitude. At other instants, the procedure repeats. It also repeats for *Dev*-sets corresponding to the longitude.

As the number of hurricanes and their duration are finite, the range of deviation values in a *Dev*-set is finite as well. Therefore, it is always possible to specify at least a single finite interval, which includes all *Dev*-values. Based on this observation, one can say that all available evidence (*Dev*-values) supports the proposition that the deviation of the latitude (computed with a given climate model using a given grid and numerical procedure) from the corresponding observational value is likely to fall inside this finite interval. Obviously, our proposition that evidence supports this specific interval is subjective and corresponds to the available database. More observational data could possibly increase the size of this interval.

The single interval supported by evidence is not very informative. Different uncertainty sources can favor different ranges of *Dev*-values. To study the distribution of *Dev*-values (*Dev*-distribution), the single interval is divided into subintervals of uniform size ΔDev, which we call the deviation step. (We assume the subintervals are of uniform size for the sake of simplicity). Each *Dev*-value unambiguously supports one of the subintervals. If the available *Dev*-values are few, or if ΔDev is small, the *Dev*-distribution will be scattered: there may be no pronounced maximum, and unsupported subintervals (*Dev*-intervals) may alternate with supported ones. As an example of the scattered distribution, Figure 1 shows the *Dev*-distribution for the latitude at $t_6 = 60h$ obtained with the $\Delta Dev = 1^o$ (ECMRWF model, 2001).

A scattered deviation distribution yields little if any useful information. A deviation distribution over the single interval and one scattered over several subintervals are two limits of possible *Dev*-distributions, which are not very informative. We observe that for the purpose of the present work, the most useful *Dev*-distribution would be one that is of the concave type, i.e., with one subinterval with maximum evidence support (more *Dev*-values fall inside this subinterval) and with the evidence to support subintervals on both sides of this subinterval monotonically decreasing. Subintervals with nonzero support are focal elements of the *Dev*-distribution and the set of all of them constitutes its core.

A concave *Dev*-distribution for a given *Dev*-set can be obtained by increasing the size of the *Dev*-step. For instance, if at $\Delta Dev = 1^o$ (the smallest *Dev*-step considered in this study) the *Dev*-distribution is scattered as the one shown in Fig.1, the next step is to redistribute the *Dev*-values over larger subintervals: $\Delta Dev = 2^o$,.. etc. until a concave distribution is

obtained. For instance, the *Dev*-distribution of the concave type for the latitude at $t_6 = 60h$ (ECMRWF model, 2001) can be obtained at $\Delta Dev = 2^o$ (see Fig. 2). Notice, that constructing a concave *Dev*-distribution, we consider fluctuations in the number of *Dev*-values supporting a single *Dev*-subinterval as a noise, if this number falls in the range between zero and 5% of the maximum of the *Dev*-distribution. As such, the fluctuations do not influence our choice of the *Dev*-step.

Although there is no guarantee that for any engineering problem there exists such a ΔDev that allows one to construct the *Dev*-distribution with the desirable property of a concave shape, it turned out to be the case for the hurricane forecasts.

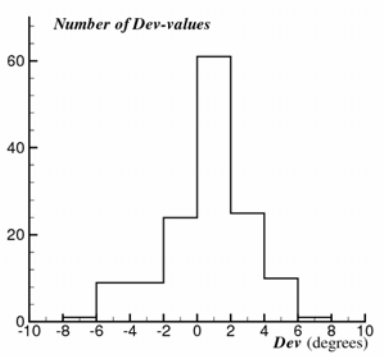


FIGURE 2. The concave *Dev*-distribution for the latitude at $t_6 = 60h$ obtained with $\Delta Dev = 2^o$. (ECMRWF model, 2001).

A concave *Dev*-distribution is then used to build an *m*-function. We define the BPA for each subinterval as the ratio of the number *n* of *Dev*-values falling inside the subinterval to the total number *N* of *Dev*-values in the *Dev*-set:

$$m(\Delta_l Dev) = n_l / N \ , \qquad (6)$$

where l is the index over focal elements of the *Dev*-distribution. In this case, because all subintervals are disjoint and there is no ambiguity in how evidence supports different subintervals, the BPA for each subinterval is equal to the degree of belief and the degree of plausibility (see expressions (2)-(3))

$$m(\Delta_l Dev) = Bel(\Delta_l Dev) = Pl(\Delta_l Dev) \ . \qquad (7)$$

In deviation distributions constructed in such a manner, the subinterval with the maximum support shows how far the uncertainties and errors in the computational procedure (which includes model uncertainty, grid resolution, observation error, etc.) will likely force the forecast results to deviate from reality (represented by observational data). Obviously, the most favorable scenario would be the one where the most supported subinterval includes the zero *Dev*-value. Another feature of a *Dev*-distribution to be considered is the size of its core. The smaller this size, the more focused is the combined contribution of uncertainty sources and better is the accuracy of the forecast. The size of ΔDev indicates whether evidence supports one subinterval over others. The smaller ΔDev one can choose without compromising the properties of the *Dev*-distribution, the better the accuracy of predictions that can be achieved, as will be shown in the following sections.

All three characteristics – the location of the maximum, the size of ΔDev and the total range of *Dev*-values -- of *Dev*-distributions depend on the forecast instant t_i, quantity (latitude/longitude), year, and model. For example, Tables 2-3 illustrate the dependence of the *Dev*-step on these parameters. Thus, these three characteristics can be used to compare, for instance, the accuracy of forecasts produced by different models and the effectiveness of changes in a model and computational procedure.

Year	1998		1999	
t_i, *h*	Latitude	Longitude	Latitude	Longitude
0	1	1	2	1
12	3	2	2	2
24	2	3	2	2
36	3	2	3	2
48	3	4	2	2
60	4	5	2	3
72	2	3	3	7
84	4	5	2	3
96	8	6	3	5
108	12	5	3	6
120	4	4	4	7

TABLE 2. The size of ΔDev for the ECMRWF model

Year	1998		1999	
t_i , h	Latitude	Longitude	Latitude	Longitude
0	2	2	1	1
12	4	2	1	1
24	4	2	1	2
36	5	2	1	3
48	5	2	4	7
60	7	3	2	3
72	9	9	7	5
84	5	4	3	3
96	9	13	3	5
108	7	4	5	6
120	8	5	4	8

TABLE 3. The size of ΔDev for the NOGAPS model

4. APPLICATIONS OF M-FUNCTIONS

In this section we will discuss how *m*-functions can be used

- to study variation of the model accuracy with the forecast duration;
- to evaluate the effectiveness of annual modifications in a model;
- to compare the performance of different models (here ECMRWF and NOGAPS models) in the years 1998-2001.

4.1 Accuracy Variation with the Forecast Duration

With each time step, the accuracy of hurricane path forecasts drops. This dynamic does not depend on the year of the forecast, the climate model, or the quantity (latitude/longitude) considered, and is quantitatively represented by the increase in the size of the *Dev*-step (see, for example, Tables 2-3). This tendency is predominantly the result of the decrease in the model's capability to correctly describe the physics of hurricanes. One could argue that the increase in the size of the *Dev*-step could also be caused by the decrease in populations of the *Dev*-sets observed with each consequent time step (see Table 1). Our study showed, however, that one could diminish fluctuations in ΔDev by increasing the size of populations, but not overcome this tendency.

Obviously, a preferred model would have the smallest and least divergent *Dev*-step for both spatial coordinates. The following two sections discuss the application of the *Dev*-step for comparing the performance of the same model after annual modifications and the performance of different models during the same year.

4.2 Effectiveness of annual model modifications

Tables 2 and 4 show the variation in the size of *Dev*-step depending on the year (1998-2001) and the forecast duration for the ECMRWF model. Again, the size of the population of a *Dev*-set is not the main reason for variations in ΔDev. For instance, at $t_1 = 0h$, populations are the largest for any year, model and quantity. However, ΔDev at $t_1 = 0h$ is not necessarily the smallest one (one degree is the smallest *Dev*-step in this study). For instance, the size of the population of the *Dev*-set for the ECMRWF model at $t_1 = 0h$ in the year 2001 is equal to 256 and at $t_4 = 36h$ (the same year) is 184. Yet, the *Dev*-step for the latitude is smaller at $t_4 = 36h$. As the other example, compare at $t_1 = 0h$ the *Dev*-steps for the latitude ($\Delta Dev = 2^o$) and the longitude ($\Delta Dev = 3^o$) obtained for the ECMRWF model in the year 2000 with the corresponding values of the year 1998 (Table 2): $\Delta Dev = 1^o$ for both quantities. The population of the *Dev*-set in 2000 is larger (281) than that of the corresponding *Dev*-set in 1998 (225). In other words, as we did in the previous section, we relate variation in the *Dev*-steps to annual modifications in a model rather than to differences in

Year	2000		2001	
t_i , h	Latitude	Longitude	Latitude	Longitude
0	2	3	2	1
12	2	2	2	1
24	2	2	3	2
36	2	2	1	2
48	2	2	2	2
60	2	3	2	3
72	2	2	2	4
84	2	2	3	2
96	2	3	3	5
108	2	6	3	6
120	3	4	4	4

TABLE 4. The size of ΔDev for the ECMRWF model

population sizes.

Tables 2 and 4 show that annual modifications introduced in the model do not always have a positive effect on the prediction accuracy. There was definite improvement in the model performance in the year 1999 in comparison with the performance of the model version of 1998. Modifications in the version of 2000 caused the lost of prediction accuracy at the beginning of forecasts, but the performance of this model version is the most consistent at different forecast instants than the three other model versions considered in this study. The 2001 version does not demonstrate significant improvement in model performance. Notice also that performance of different model versions at $t_{11} = 120h$ does not practically change from year to year, as if there were no modifications at all.

As we see, variation in the size of the Dev-step can be used to quantitatively evaluate the effect of model modification. Comparison of model versions of different years should preferably be performed using the observational database of one year to eliminate the influence of the different population sizes of the *Dev*-sets. Unfortunately, we did not have such an opportunity in this study. Developers of climate models can, however, use this technique in their study to eliminate costly model modifications with little or negative effect.

Tables 3 and 5 illustrate the effect of annual modifications in the NOGAPS model.

Year	2000		2001	
t_i , h	Latitude	Longitude	Latitude	Longitude
0	1	1	1	1
12	1	2	2	1
24	3	2	4	1
36	11	2	2	1
48	4	3	3	4
60	2	2	2	4
72	3	5	5	2
84	3	5	6	2
96	2	6	8	3
108	4	9	9	4
120	7	3	5	3

TABLE 5. The size of ΔDev for the NOGAPS model

In summary, variations in ΔDev demonstrate that model modifications do not always improve the predictions accuracy. They rarely decrease the accuracy though.

4.3 Comparison of ECMRWF and NOGAPS models

If one is interested in evaluation of the overall model performance in a given year, without distinguishing between forecast instants, Tables 2-5 can be presented in a summarized form:

	1998	1999	2000	2001
ECMRWF	40	40	31	32
NOGAPS	48	44	40	26

TABLE 6. Sums of ΔDev from $t_1 = 0h$ to $t_{11} = 120h$ for the latitude.

	1998	1999	2000	2001
ECMRWF	46	28	23	27
NOGAPS	65	32	41	47

TABLE 7. Sums of ΔDev from $t_1 = 0h$ to $t_{11} = 120h$ for the longitude.

Tables 6-7 serve two purposes. They can be used to evaluate the total effectiveness of annual model modifications or to compare the performance of different models. In the first case, these tables confirm the conclusion we have made in the previous section that the ECMRWF model of the year 2000 performs better than the other three versions for the model. As for the NOGAPS model, versions for the years 1999 and 2001 are rather similar in accuracy. Versions of the years 1998 and 2000 produce less accurate results.

If one compares the accuracy of the two models, it is clear from Tables 6-7 that the ECMRWF model consistently performs better than the NOGAPS model.

Similar analysis can be conducted for two other characteristics of the *m*-function – the size of its core and the position of its maximum. However, we do not discuss these results here.

5. UNCERTAINTY QUANTIFICATION IN HURRICANE PATH FORECASTS

In this section, we explore the possibility of using *m*-functions to quantify, and possibly, improve the accuracy of hurricane path forecasts in situations where no observational data is available. The procedure we developed relies on the results of computations with climate models and *Dev*-distributions constructed for these models using the observational database of previous hurricanes. In the procedure we do not choose between various climate models. Instead, we fuse the information they provide. The steps of the procedure are highlighted below. The detailed description of each step is given in the sections that follow.

Step 1: Dev-distributions are constructed for all models included in the forecast of a given hurricane. The accuracy of each model is characterized by *Dev*-distributions different for the latitude and the longitude, and the forecast instants. The observational database of previous hurricanes is used for that. Ideally, these would be the hurricanes of previous years. However, annual modifications in any model result in fact in a new model under the same name. Therefore, information on the model accuracy collected in the previous years cannot be used in the year of interest. As a result, the path of the first hurricane of a season cannot be predicted by using the procedure we describe in this paper. Also, the accuracy of predictions will increase with increasing the size of the observational database. In the future, we plan to conduct a more thorough study on the evolution of forecast accuracy depending on the size of the database. Notice that the problem of the size of the observational database is common for any approach requiring the training phase (Kumar et al., 2003). Difficulties could be overcome though if modelers would present the *Dev*-distributions constructed using the observational database of previous years along with a new model version.

Step 2: Dev-distributions are transformed in the *m*-functions.

Step 3: The *m*-functions characterizing the accuracy of a climate model in predicting the latitude (longitude) at different forecast instants are combined with the prediction made using the model for a hurricane for which observational data are not available. The result for a model at each instant of the forecast is a grid centered on the model prediction of the hurricane position. For each grid interval in latitude and longitude directions, the BPA is prescribed. Its value reflects our belief that the latitude (longitude) of the real hurricane position falls exactly within that interval.

Step 4: Then, the resultant predictions of several models are fused using Dempster's rule (4) of evidence theory to create a new prediction. Fusion is performed independently for each coordinate and at each instant of the forecast. The final prediction is again a grid (different at different instants), where intervals in spatial directions are characterized by BPAs obtained from (4). The BPA of an entire grid cell is the product of BPAs of the corresponding grid intervals in the latitude and longitude directions. The grid cell with the highest BPA is also the one with the highest degree of belief, and as such is the most likely candidate for hurricane's position.

Step 5: At each forecast instant, the grid cell with the highest degree of belief is extracted. Such cells form a hurricane path, which we call the swath of maximum degree of belief. The swath of maximum degree of belief is supposed to be more reliable than individual model forecasts, and ideally, coincides with observational data, if they are available.

In this study, the data for hurricanes of the year 2000 are used to evaluate the approach we developed to fuse forecasts produced by two models: the ECMRWF model and the NOGAPS model. The procedure is described for this case. Other models will be included in future studies. For convenience, we numerate hurricanes in accordance with the order of their occurrence during the year and a region (South, East, West). For instance, 4S stands for the fourth hurricane in the available database, which happened in the South region of the Pacific Ocean. Observational data for three hurricanes from the

South Pacific region, three hurricanes from the East Pacific region, and six hurricanes from the West Pacific region are used solely for the evaluation of the quality of predictions obtained with the new technique. The data for other hurricanes of the year 2000 are used to quantify the model uncertainty at Step 1.

5.1 Step 1

The procedure starts from building *Dev*-distributions characterizing the accuracy of each model in predicting hurricane paths. Since we use Dempster's rule to fuse information from different sources (forecasts produced by different climate models), it is required that the degrees of belief be based on independent sources of evidence. Independence of evidence sources is important, but its definition is highly subjective (Shafer, 1976b). As we are working with *Dev*-distributions, it stands to reason to assume that the *Dev*-distributions are independent of one another if they are constructed using the results of forecasts produced by different climate models and different observational data. In the current study, it means that the observational databases used to build *Dev*-distributions related to the NOGAPS model and the ECMRWF model should be different. We will explore in detail the possibility of relaxing this demand in the future. In this study, however, we simply divide the data for hurricanes of the year 2000 in two groups. Then, for each model, we obtain two independent sets of *Dev*-distributions, with one set corresponding to one part of the observational data and another set corresponding to the other part of the data. That is, for each quantity (latitude or longitude) at each forecast instant, we have two independent *Dev*-distributions for a given model. To simplify the discussion, we denote the set of *Dev*-distributions corresponding to the ECMRWF model and the part one of the observational data as E1; N1 denotes the set of *Dev*-distributions corresponding to the NOGAPS model and the same observational data. E2 and N2 denote the sets for the ECMRWF and NOGAPS models, respectively, obtained using another part of observational data.

Section 3 provides the details on the construction of *Dev*-distributions.

5.2 Step 2

The *m*-functions are determined by *Dev*-distributions. The number of *Dev*-values, which constitutes a *Dev*-distribution, depends on the forecast instant, the model, and the part of observational data it corresponds. For instance, at $t_1 = 0h$, the *Dev*-distributions for the latitude and longitude of the E1-set are based on 154 observational data; those of the E2-set are based on 127 data. Corresponding numbers for N1- and N2-sets are 280 and 261, respectively. Intuitively, the more observational data used, the more confidence we have in a *Dev*-distribution. This should be reflected in the degrees of belief we assign to subintervals of corresponding *m*-functions.

For example, let us consider two *Dev*-distributions -- Dev_1 and Dev_2 -- that are constructed based on observational datasets N_1 and N_2, respectively. Assume that $N_1 > N_2$. Then, the BPAs for subintervals of the Dev_1-distribution calculated by expression (6) do not change. In other words, the total belief in the distribution built using the largest number of available experimental data does not change. For the Dev_2-distribution, the BPAs have to be recalculated though:

$$m(\Delta_i Dev_2) = n_i / N_1 , \qquad (8)$$

where i is the index over subintervals with nonzero support of the Dev_2-distribution and $\sum_i n_i = N_2$. The sum of BPAs determined by Equation (8) is less than 1:

$$\sum_i m(\Delta_i Dev) = \frac{N_2}{N_1} < 1 .$$

To satisfy condition (1), we assign a BPA equal to $(N_1 - N_2)/N_1$ not to any individual subinterval specifically, but to the set of all possible *Dev*-values and call it uncommitted belief. It reflects the fact that if

we had additional $N_1 - N_2$ experimental data, we would not know which *Dev*-subintervals they would support.

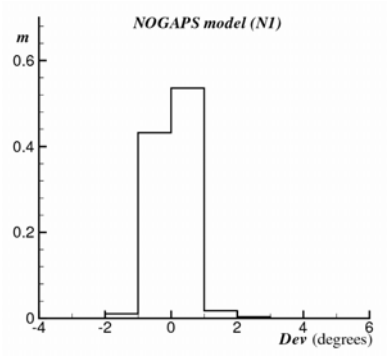
In our case, we have four *Dev*-distributions based on different numbers of *Dev*-values at each forecast instant. For the *Dev*-distribution based on the largest number of Dev-values, we use expression (6) to calculate BPAs for its individual subintervals. Expression (8) is used to calculate BPAs for individual subintervals of the other three *Dev*-distributions. For instance, at $t_1 = 0h$ the *m*-functions for both the latitude and the longitude corresponding to the Dev-distributions of the N1-set are constructed using expression (6). The *m*-functions for the other three *Dev*-distributions are constructed using (8). A BPA equal to $(280-154)/280 = 126/280$ is assigned as the uncommitted belief to the *Dev*-distributions of the E1-set. Uncommitted beliefs for the *Dev*-distributions of E2- and N2-sets are $153/280$ and $19/280$, respectively. The BPAs committed to individual subintervals of all four *Dev*-distributions for the latitude at $t_1 = 0h$ are shown in Fig. 3. Uncommitted belief is not shown in Figs. 3b-3d.

Notice that since we consider different forecast instants as independent in this procedure, we do not introduce in the *m*-functions the uncertainty associated with different numbers of *Dev*-values available at different forecast instants.

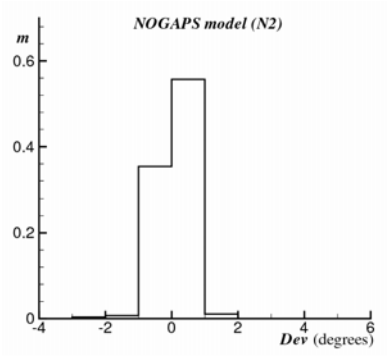
5.3 Step 3

Expression (5) can be used for prediction in the following manner. If one knows the model data (L^m) for the position of a given hurricane at a given instant and knows the deviation values corresponding to the latitude and the longitude at this instant, one can try to define the "true" hurricane position, that is, the position, which would coincide with the observational data if available. There is no guarantee, but it is our belief that the hurricane position found in such a way would better reflect reality. Thus, we rewrite expression (5) in the following way:

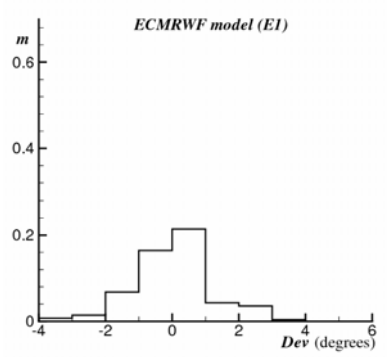
$$L^p = Dev + L^m, \qquad (9)$$



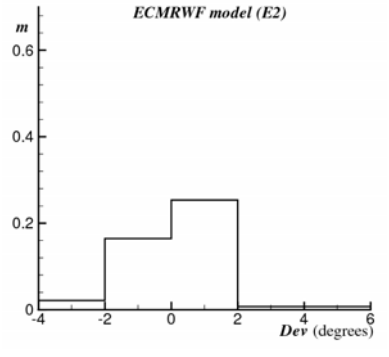
a)



b)



c)



d)

FIGURE 3. The *m*-functions for the latitude at $t_1 = 0h$.

where L^p is the predicted value of the latitude or the longitude. In reality, we do not know the exact *Dev*-values. We do know, however, the *Dev*-distributions built in Step 1, which provide information about which *Dev*-subintervals have nonzero BPA at a given forecast instant. Corresponding BPAs of *Dev*-distributions are determined is Step 2. So, instead of one value for a spatial coordinate, which would be given by expression (9), we have information on how different latitude/longitude intervals would be supported by evidence, which in this case is given by BPAs for *Dev*-subintervals. If L^m is the value of the latitude (longitude) calculated by a climate model for a given hurricane at a given forecast instant, and $\Delta_i Dev$ are subintervals of the corresponding *Dev*-distribution obtained by analyzing previous forecasts produced by this model, then we determine the supported latitude (longitude) intervals as

$$\Delta_i L^p = L^m + \Delta_i Dev \,. \qquad (10)$$

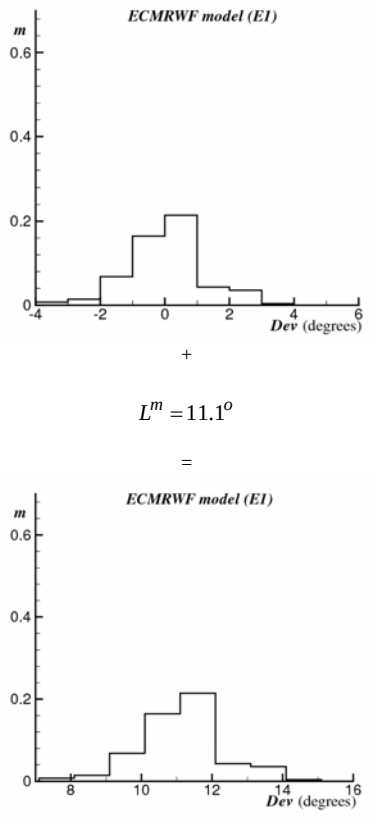


FIGURE 4. Illustration to Step 3

The BPAs of *Dev*-subintervals are directly reassigned to corresponding latitude (longitude) intervals as

$$m(\Delta_i L^p) = m(\Delta_i Dev) \,. \qquad (11)$$

As an example, Figure 4 shows BPAs for the individual latitude intervals at $t_1 = 0h$. These BPAs correspond to the *m*-function obtained by combining the *m*-function shown in Fig. 3c (the ECMRWF model, E1 set) and the model prediction of the latitude value ($L^m = 11.1^o$) for a hurricane for which we assume that no observational data is available.

This procedure explains why the single maximum is a desirable property of a *Dev*-distribution. For a given hurricane at a given forecast moment, there could be only one value of the latitude (longitude), which characterizes the hurricane position. Correspondingly, there could be only one interval that includes this value. Thus, evidence should favor one interval over others to avoid contradiction.

To summarize, the result from Step 3 for each model at each forecast instant is a grid centered on the model prediction of the hurricane position. For each grid interval in latitude and longitude directions, the BPA is prescribed. Since individual intervals of *m*-functions do not intersect, the BPA of each interval coincides with the total belief assigned to the interval. Its value reflects our belief that the latitude (longitude) of the real hurricane position falls exactly within the interval. In fact, for each model at a given instant, there are two, not one, equally likely grids with assigned beliefs. The two grids correspond to two different sets of *m*-functions existing for the same model, e.g., sets E1 and E2 for the ECMRWF model.

5.4 Step 4

In this step, the resultant predictions of two models are fused using Dempster's rule (4) of evidence theory to create a new prediction. Fusion is performed independently at each instant of the forecast for each coordinate. First, we fuse the *m*-functions of the ECMRWF model belonging to the E1 set with the *m*-functions of the NOGAPS model belonging to the N2 set. We call the resultant prediction the E1N2

solution. By combining *m*-functions from different sets of different models we satisfy the requirement of Dempster's rule for independence of evidence sources (for more detail see the discussion of Step 1). Another requirement of Dempster's rule is that evidence from different sources should not strictly contradict each other. This requirement is implicitly satisfied in this problem: areas of possible values of the latitude (longitude) corresponding to different climate models overlap at any forecast instant. In problems where supported areas do not overlap, other fusing techniques should be considered instead of Dempster's rule.

Let us demonstrate how Dempster's rule works in a simple case. Assume that at a given forecast instant, the *m*-function for the latitude of the ECMRWF model consists only of the single interval, ΔL_{E1}, with the BPA $m_1(\Delta L_{E1}) = s_1$ and with the rest of belief assigned to the set of all possible latitude values $m_1(\mathbf{L}) = 1 - s_1$ (uncommitted belief). At the same instant, the *m*-function of the NOGAPS model also consists of one interval ΔL_{N2} with the BPA $m_2(\Delta L_{N2}) = s_2$ and with the uncommitted belief $m_2(\mathbf{L}) = 1 - s_2$. The result of fusing two basic probability assignment functions, m_1 and m_2, can be presented as a table and is shown in Fig. 5.

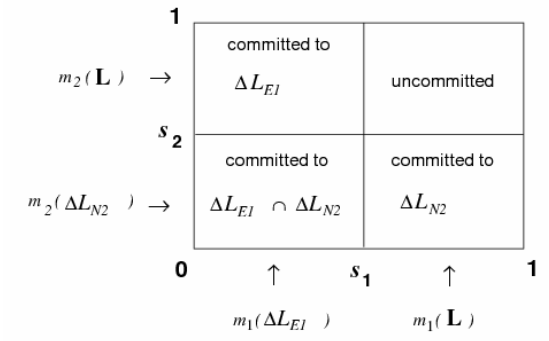


FIGURE 5. Example of Dempster's rule application

The entire table represents our total belief. Vertical strips are associated with BPAs of latitude intervals provided by the ECMRWF model and horizontal strips are associated with BPAs of latitude intervals provided by the NOGAPS model. The intersection of these two strips has measure, $m_1(\Delta_i L) \cdot m_2(\Delta_j L)$, where *i* and *j* are indices over intervals of functions m_1 and m_2, respectively. In our case, $i = j = 1$. The BPA of the intersection of two intervals is

$$m(\Delta L_{E1} \cap \Delta L_{N2} \neq \varnothing) = m_1(\Delta L_{E1}) \cdot m_2(\Delta L_{N2}) .$$

If each function, m_1 and m_2, assigns nonzero BPAs for several intervals, then an interval of m_1 can intersect with more than one interval of m_2. In this case, the BPA of the interval is the sum of measures of all related intersections. As Fig. 5 demonstrates, the uncommitted belief $m_2(\mathbf{L})$ contributes to the BPA of the ΔL_{E1}-interval, and in a similar manner, the uncommitted belief $m_1(\mathbf{L})$ contributes to the BPA of the ΔL_{N2}-interval. The product of uncommitted beliefs does not relate to any specific interval. If two intervals do not intersect, the measure,

$$m(\Delta L_{E1} \cap \Delta L_{N2} = \varnothing) = m_1(\Delta L_{E1}) \cdot m_2(\Delta L_{N2}) ,$$

should be deduced from the total belief. Then, the BPAs for intersecting intervals should be renormalized accordingly. This is how expression (4) for Dempster's rule is derived.

No area in Fig. 5 should be deduced from the total belief. So, renormalization is not required. Therefore, the basic probability assignment function *m,* which corresponds to Fig. 5, is defined as

$$m(\Delta L_{FUS}) = \begin{cases} s_1 \cdot s_2, \text{ if } \Delta L_{FUS} = \Delta L_{E1} \cap \Delta L_{N2}; \\ s_1 \cdot (1 - s_2), \text{ if } \Delta L_{FUS} = \Delta L_{E1}; \\ s_2 \cdot (1 - s_1), \text{ if } \Delta L_{FUS} = \Delta L_{N2}; \\ (1 - s_1)(1 - s_2), \text{ if } \Delta L_{FUS} = \mathbf{L}; \\ 0, \text{ if } \Delta L_{FUS} \text{ is any other interval} \\ \quad \text{and } \Delta L_{FUS} < \mathbf{L}. \end{cases} \qquad (12)$$

The corresponding *Bel*-function can be obtained from expression (2); plausibility functions are not used in the current study.

It is easy to see that fusing even the simplest *m*-functions (defined over a single interval) results in a complex *m*-function defined over several intervals of different size. Keeping in mind that i) *m*-functions, as those built in Steps 2-3, include several intervals and

ii) intervals of two *m*-functions intersect in different ways at different forecast instants, we decided to coarsen the set of intervals after fusing in order to avoid unnecessary complexity. The coarsening means that we calculate BPAs only for the set of intervals associated with one of the *m*-functions used in fusing. This set has a minimum deviation step ΔDev. If the deviation steps of two *m*-functions are of the same size, we choose the set of intervals with the highest degree of belief. BPAs for the chosen set are obtained by Dempster's rule (expression (4)). As the intervals of this set are disjoint, the total belief that the "true" velocity value is contained in an interval is equal to the BPA of that interval.

The final prediction is again a grid (different at different instants), where intervals in spatial directions are characterized by the BPAs. BPA of an entire grid cell is the product of BPAs of the corresponding grid intervals in the latitude and longitude directions:

$$m(\Delta_i Lat \times \Delta_j Lon) = m_1(\Delta_i Lat) \times m_2(\Delta_j Lon)$$

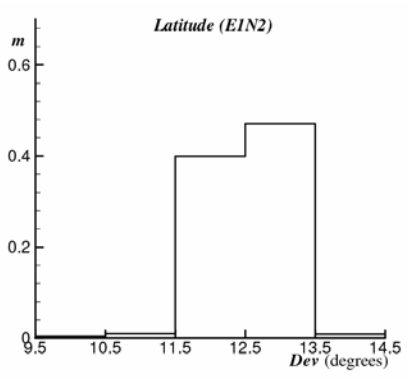
(*i, j* are indices over intervals of functions m_1 and m_2 corresponding to the latitude (Lat) and the longitude (Lon), respectively). The grid cell with the highest BPA is also the one with the highest degree of belief, and as such is the most likely candidate for the hurricane position.

To illustrate the procedure, let us combine two *m*-functions for the latitude belonging to E1 and N2 sets (Figs. 3b,c) and corresponding *m*-functions for the longitude (not shown here) with the predictions made with NOGAPS and ECMRWF models at $t_1 = 0h$ for the position of the same hurricane:

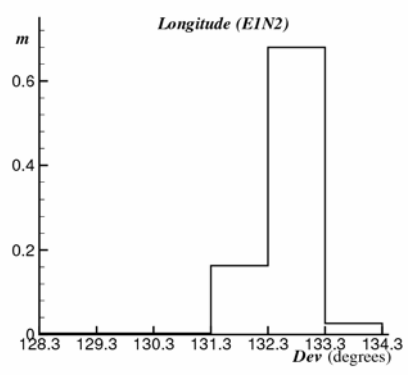
	Latitude	Longitude
ECMRWF	11.1	134.5
NOGAPS	12.5	132.3

TABLE 6. Models predictions of the hurricane position at $t_1 = 0h$

The *m*-function resulting from fusing two *m*-functions for the latitude obtained in Step 3 is shown in Fig. 6a; the resultant *m*-function for the longitude is given in Fig. 6b. Figure 7a shows the distribution of belief over all grid cells in this case. Different colors correspond to different degrees of belief. There is no belief that the hurricane position would be outside the colored area.



a)

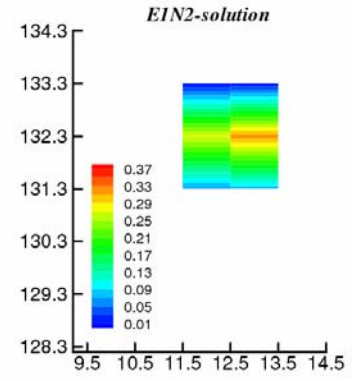


b)

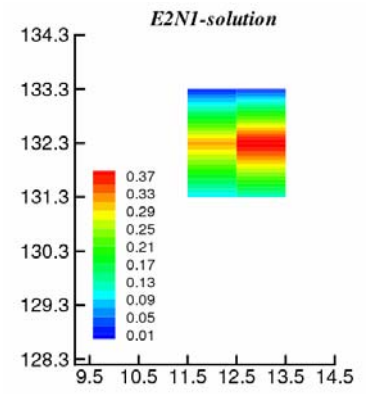
FIGURE 6. The *m*-functions for the latitude (a) and the longitude (b), which result from the application of Dempster's rule to *m*-functions shown in Figs. 3b, 3c and corresponding *m*-functions for the longitude combined with the model data of Table 6.

In a similar manner, we fuse the *m*-functions of the ECMRWF model belonging to the E2 set with the *m*-functions of the NOGAPS model belonging to the N1 set and call the resultant prediction the E2N1solution. The E2N1 solution corresponding to the E1N2 solutions shown in Fig. 7a, is given in Fig. 7b.

Two solutions – E1N2 and E2N1 – are equally likely. Therefore, we average these two solutions to produce a single forecast. In this step, two solutions R1 and R2 are averaged. The intervals of the two solutions are combined individually for each coordinate and independently at each forecast instant. In order to resolve the mismatches in size and



a)



b)

FIGURE 7. Solutions E1N2 (a) and E2N1 (b) corresponding to *m*-functions shown in Fig. 3 and the model data of Table 6

location of the intervals of the two solutions, we choose the most refined interval set (with a minimum ΔL) to increase the accuracy of predictions. Then, the other solution is projected onto the chosen set. The procedure is described below in detail.

Let us assume that at a given instant the grid of the E1N2 solution in the latitude direction is more refined than the grid of the E2N1 solution in the same direction. Therefore, we will project the solution E2N1 for the latitude onto the latitude intervals of the solution E1N2. At a given forecast instant, m_{11} is the BPA of the latitude interval $[L_{11}, L_{12}]$ of the solution E1N2; m_{21} is the BPA of the latitude interval $[L_{21}, L_{22}]$ and m_{22} is the BPA of the velocity interval $[L_{22}, L_{23}]$ of the solution R2. The interval $[L_{11}, L_{12}]$ intersects both intervals of the solution R2 as shown in Fig. 8.

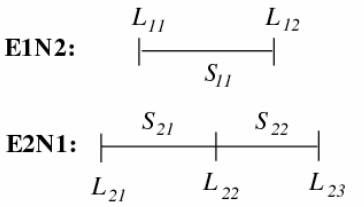


FIGURE 8. Intersecting intervals of two solutions

Then, the BPA of the interval $[L_{11}, L_{12}]$ for the averaged solution is calculated by the formula

$$m = \frac{1}{2}\left(m_{11} + m_{21}\frac{m_{22} - m_{11}}{m_{22} - m_{21}} + m_{22}\frac{m_{12} - m_{22}}{m_{23} - m_{22}} \right),$$

which takes into account m_{21} and m_{22} with appropriate weights. Weights are determined by the fraction of an interval that overlaps with $[L_{11}, L_{12}]$. For other types of interval overlapping, a similar approach for calculating the averaged interval BPA should be applied. Figure 9 shows the result of averaging E1N2 and E2N1 solutions from Fig. 7.

The averaging technique used in this paper is one of the simplest and is well suited to the present study.

5.5 Step 5

Figure 9 gives an example of the prediction resultant from Steps 1-4 at a single forecast instant. The complete forecast of a hurricane consists of predictions at several instants. The number of instants is determined in our procedure by the number of instants at which the data of all models included in the forecast are available. An example of the complete

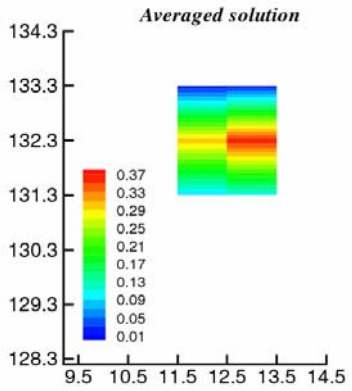


FIGURE 9. Averaged solution corresponding to E1N2 and E2N1 solutions shown in Fig. 7.

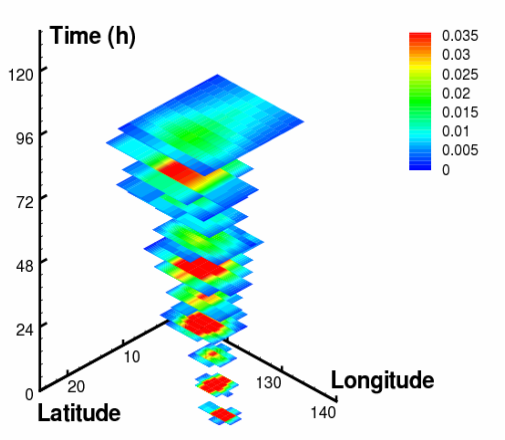


FIGURE 10. Complete forecast for the hurricane 1W

forecast is given in Fig. 10. In this forecast, the prediction at $t_1 = 0h$ coincides with the one shown in Fig. 9. Different colors reflect different degrees of belief. For better visualization, areas with low or zero degrees of belief are cut. This is a forecast for the 1W hurricane started on May 6, 2000 at 12.00 p.m. in the West region of the Pacific Ocean.

To serve better practical purposes, a forecast can be reduced to a single grid cell with the highest degree of belief at each forecast instant.

Such cells form a hurricane path, which we call the path of maximum degree of belief. The path of maximum degree of belief is the most probable candidate to include the "true" hurricane path.

The path shown in Fig. 11 is extracted from the complete forecast of the hurricane 1W given in Fig. 10. In the figure, the path is compared with the predictions made with each climate model individually (white circles correspond to the ECMRWF model prediction and grey circles correspond to the NOGAPS model prediction). We also compare the path with the ensemble averaged data (grey squares) (the ensemble consists of the two models). The observational data (black circles) are shown in the figure to assess the quality of the prediction.

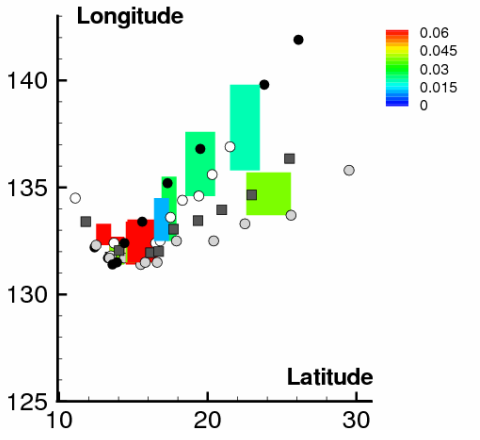


FIGURE 11. The path of maximum degree of belief for the hurricane 1W

One can see that the NOGAPS model prediction in this case is far from observational data, whereas the ECMRWF model is in good agreement with observations. Our approach combines the results of both models, and yet, our prediction includes observational data in the path of maximum degree of belief at all forecast instants up to 120*h*, except $t_{10} = 108h$. These results show a good potential of the approach to correctly "weight" contributions from different sources. Comparison with the ensemble averaged data also favors our approach. Also, in contrast to the predictions made by individual models or by their average, our approach produces not just a single line, the accuracy of which cannot be estimated in the absence of observational data, but zones with well-defined degrees of belief. This is an obvious advantage of the present method.

Notice that the size of the grid cell with the maximum degree of belief increases with the forecast duration, and the degree of belief assigned to the cell diminishes. This is the consequence of the drop in models accuracy as we discussed above in Section 4.

In the following section we provide additional examples of the application of the approach we described above.

6. EXAMPLES OF PREDICTIONS

Figures below show forecasts for four different hurricanes occurred in the year 2000 in the different regions of the Pacific Ocean. Denotations are the same as in Fig. 11. The observational paths of these hurricanes are complex and it makes them difficult to predict. Yet, our approach catches hurricanes trends well. Its performance is better and more consistent

than the performance of individual models and their average.

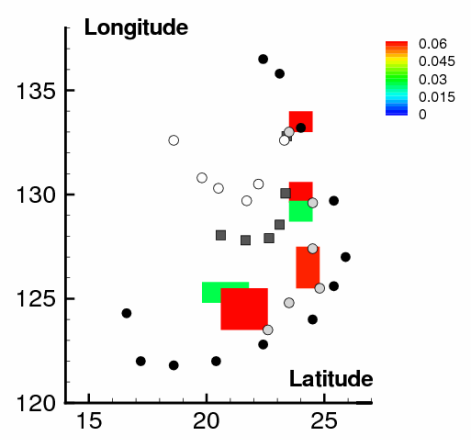


FIGURE 12. Hurricane 4S (start time: 01/06, 12.00 p.m.)

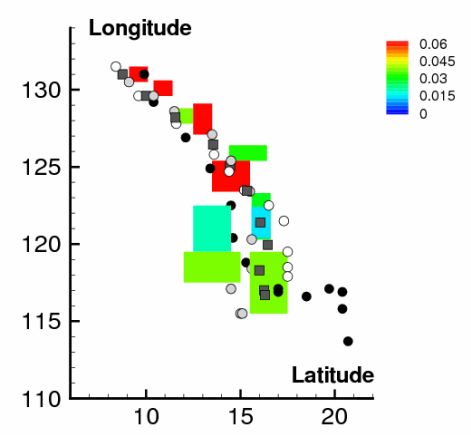


FIGURE 13. Hurricane 3E (start time: 06/21, 12.00 p.m.)

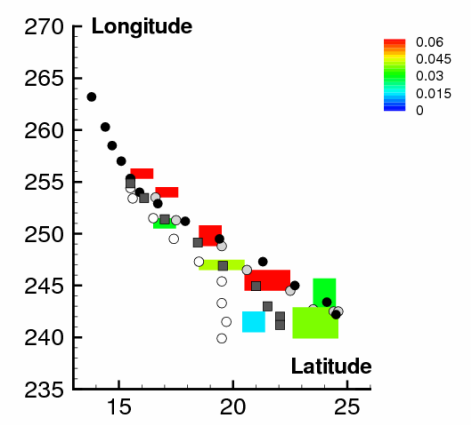


FIGURE 14. Hurricane 24W (start time: 09/07, 12.00 p.m.)

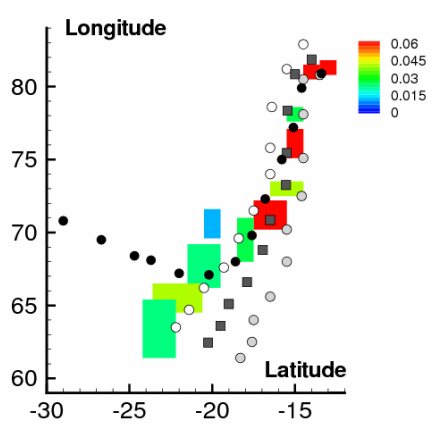


FIGURE 15. Hurricane 31W (start time: 10/31, 12.00 p.m.)

7. CONCLUSIONS

In this paper, we describe a new technique to quantify the accuracy of climate models in predicting the hurricane path using the available database. The technique proves to be effective for studying the evolution of the model accuracy with a forecast duration and the effectiveness of annual model modifications, and for comparing the performance of different models. Results of testing the technique using two climate models: the ECMRWF model and the NOGAPS model and observational database for hurricanes of the years 1998-2001 in the Pacific Ocean are shown.

We also suggest an approach to quantify and improve the accuracy of hurricane path forecasts in situations where no observational data is available. The approach relies on the mathematical tools of evidence theory, which are customized here for application to total uncertainty in simulations.

Application of this approach to the hurricane forecasts of the year 2000 has provided encouraging results. In the future, we plan to further improve the accuracy of predictions produced by the approach. One of the possibilities is to increase the number of climate models included in the forecast. The other possibility is to apply to the solution obtained at Step 5, the smoothing procedure described in Poroseva et al. (2005) prior extracting the path of maximum degree of belief.

Acknowledgments

Authors express their gratitude to Professor T. N. Krishnamurti and Dr. V. Kumar (Department of Meteorology, Florida State University) for providing observational and model data necessary for this study.

References

Dubois, D. and Prade, H., *Possibility Theory: an Approach to Computerized Processing of Uncertainty*, Plenum Press, New York, 1988.

Klir, G. J. and Wierman, M. J., *Uncertainty-Based Information: elements of generalized information theory*, Heidelberg; New York: Physical-Verl., 1998.

Kumar, T. S. V. V., Krishnamurti, T. N., Fiorino, M., Nagata, M., 2003: Multimodel Superensemble Forecasting of Tropical Cyclones in the Pacific. *Monthly Weather Rev.*, **131**, 574-583.

Oberkampf, W. L., Trucano, T. G., and Hirsch, C., 2004: Verification, Validation, and Predictive Capability in Computational Engineering and Physics. *Appl. Mech. Rev.,* **57**(5), 345-384.

Poroseva, S. V., Hussaini, M. Y., and Woodruff, S. L., 2005: On improving the predictive capability of turbulence models using evidence theory. *AIAA-2005-1096*.

Sentz, K. and Ferson, S., 2002: Combination of Evidence in Dempster-Shafer Theory. *SAND2002-0835.*

Shafer, G., 1976a: *A Mathematical Theory of Evidence*. Princeton, NJ: Princeton University Press.

Shafer, G., 1976b: A Theory of Statistical Evidence. *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, edited by W. L. Harper and C. A. Hooker, Reidel, Vol. II, 365-436.

Shafer, G., 1987: Belief Functions and Possibility Measures. *The Analysis of Fuzzy Information*, edited by James C. Bezdek, CRC Press, Vol. 1, 51-84.

Shafer, G., 1990: Perspectives on the Theory and Practice of Belief Functions. *Int. J. Approximate Reasoning*, **4**, 323-362.

Williford, C. E., Krishnamurti, T. N., Torres, R. C. et al., 2003: Real-time multimodel superensemble forecasts of Atlantic tropical systems of 1999. *Monthly Weather Rev.*, **131**, 1878-1894.

Yager, R. R., Kacprzyk, J., and Fedrizzi, M. (editors), 1994: *Advances in the Dempster-Shafer Theory of Evidence,* New York: John Wiley & Sons, Inc.