J1

DETERMINING THE LIKELIHOOD OF SEVERE WEATHER BASED ON MODEL OUPUT

Stephen M. Jaye*
University of Wisconsin, Madison, WI

1. INTRODUCTION

The North American Regional Reanalysis (NARR) provides historic weather data over North America with higher spatial and temporal resolution than the global reanalysis (Mesiger et. al. 2002). This dataset can be used to provide a much more accurate assessment of the atmospheric conditions associated with each severe weather event. This can be used with the Storm Prediction Center's storm data, which is an archive of the location, type, intensity, and time of every official storm report, to create a climatology of atmospheric conditions associated with severe weather.

With this climatology, a procedure can be created to predict the likelihood of severe weather based on computer model output. This procedure is designed to use model output to first calculate the value of several parameters and use the relationships between these parameters and severe weather to produce a conditional likelihood of severe weather based on atmospheric conditions. The parameters used to create a conditional likelihood of severe weather were selected from a field of fifty possible parameters based on their relationship with severe weather, and their correlations with each other. The relationship between each parameter and all three types of severe weather were determined from the NARR and SPC storm data for the years 1979 through 2003.

This procedure will produce severe weather probabilities for all three types of severe weather, which are defined as the presence of a tornado, damaging winds of over fifty knots, and large hail of diameter 0.75" or greater. The probabilities outputted from the procedure will be similar to the SPC's convective outlooks, which define their probabilistics as the probability that a severe storm will be reported within twenty-five miles of a given point. It will also be possible to use this procedure with any model that produces ample resolution both temporally and spatially. This procedure will later be used in conjunction with the North American Regional Climate Change Assessment Program to recreate the current climate of severe weather, and then attempt to forecast the future state of severe weather climate.

* *Corresponding author address:* Stephan M. Jaye, Univ. of Wisconsin, Dept. of Atmospheric and Oceanic Sciences, Madison, WI 53705; email: smjaye@wisc.edu.

2. DATA

The two data sets used in creating the climatology of atmospheric conditions associated with severe weather are the North American Regional Reanalysis and the Storm Prediction Center's severe storm archive. The NARR dataset begins with the year 1979, which is the first year of twenty-five selected for the climatology. The NARR provides a reanalysis over North America at the special resolution of around 30 km every three hours. This provides a much more accurate representation of the actual conditions associated with a severe weather event than the global reanalysis would.

The SPC has archived severe weather data that dates back to 1955. However, since there is no NARR before 1979, the data from 1979 through 2003 was used to produce the climatology. This data includes the time and location of every severe storm event. For a storm to be included in this dataset, it must be declared official by the SPC. This quality check prevents faulty storm reports, which occur regularly, from making it into the dataset.

The biggest issue with the SPC storm data is the possibility of missed storms, especially early in the climatology. Since severe storms often frequent a region of the country that is quite sparsely populated, many storms occur that are never seen by humans, and therefore not reported. With the rising number of storm spotters and storm chasers in the last two decades, the number of storms reported each year has gone up. It is pretty certain that some storms have been missed by this dataset. The effect this has on the climatology of atmospheric conditions associated with severe storms in not known. The main effect should be conditional frequencies that are lower than the actual value, but this effect might not be that significant.

3. VARIABLE SELECTION

Certain parameters have been shown by

$$P(tornado / parameter\ value) = \frac{occurences\ of\ tornadoes\ with\ this\ paramter\ value}{total\ occurencs\ of\ this\ parameter\ value} \qquad (1)$$

recent study to be good indicators of favorable atmospheric conditions for severe weather development. These parameters are used as a mechanism in which to describe the state of the atmosphere. Fifty possible predictor variables were selected from previous studies, and forecast products. Many of the predictor variables have been known to the science for a long time, such as Convective Availible Potential Energy (CAPE) and Convective Inhibition (CIN). More recent previous studies, such as Weismen and Klemp (1982) developed specific parameters to forecast likelihood of severe weather. Other studies, such as Rasmussen and Blanchard (1998) and Thompson et. al. (2003), examined the ability for these predictors to forecast tornadoes. Forecast parameters developed in the SPC's tornado, hail, and severe indices available in the forecast tools in their website were included as well.

These fifty predictor variables were all calculated at every grid point for every file in the NARR from 1979 through 2003. This provided a value for all fifty potential predictor variables every three hours for twenty-five years. The overall probabilities of all possible values of all the predictor variables were calculated for these twenty-five years. This was used to determine the denominator in the equation below, which is based on the overall probability distribution of occurrence of each variable.

Also, using the SPC severe storm data, the value of all fifty predictor variables associated with each tornado, large hail, and damaging wind report was calculated. The value of a given parameter associated with a given storm report was taken to be the nearest spatial point in the reanalysis, and the most recent reanalysis time. Since the NARR has 30 km resolution, the nearest point is never more than 22 km, or 14 miles, away from the actual storm location. Although mesoscale weather can occur on smaller spatial scales than 22 km, the larger scale weather conditions leading to a storm is not likely differ too greatly 22 km in any direction. Using the most recent time prior to a storm ensures the environment was not represented by air that was modified by the storm itself. Some storms could be represented by atmospheric conditions as much as three hours prior to the storm. This represents a fairly small fraction of the storms, and only a fraction of these storms would be represented by atmospheric conditions that have been significantly altered between the time of the most recent reanalysis and the time of the storm's actual occurrence.

With the values of all fifty parameters associated with each storm for the 1979-2003 period, the overall probabilities of tornadoes, damaging wind, and large hail occurring for all possible values of each parameter can be calculated using Equation 1.

This calculation can be done for tornadoes, damaging wind, and large hail, and can be done with any parameter. With these calculations, charts that show the values of conditional likelihood of severe weather vs. each parameter were made. These charts were the main consideration in variable selection, as they showed how well a variable can predict severe weather on its own.

Another important consideration in variable selection was the correlations between the predictor variables. In a Bayesian framework, two variables that are highly correlated provide little to no more information than a single variable. The end result was two variables that represent the same aspect of a storm, such as CAPE and Lifted Index (LI), were not both used. Instead, the one that provided a more robust relationship, which proved to be LI, was chosen, and CAPE, which has a correlation of –0.97 with LI, was discarded.

The final consideration was the number of storms the variable actually captures. This ensures that the procedure created from these variables would actually capture the many different types of storms that produce severe weather. To ensure the procedure would do more than identify the few cases that contain exact ideal conditions, the percentage of storms with high conditional frequencies was also calculated for each possible predictor.

Table 1 shows the seven variables that were selected for use in the model procedure. The calculations made show a more robust relationship between severe weather and shear when it is scaled to the center of buoyancy of a storm rather than to an arbitrary level such as 1km, or 6km. The same is true for helicity. Scaling the parameter calculations to the center of maximum buoyancy allows these values to adjust to the specifics of a storm, which would be more likely to identify severe weather from many different types of storms.

Convective Inhibition	0-CMB wind shear	Lifted Index	W @ Level of Free Convection	0-CMB storm relative helicity	Surface Streamwise vorticity	Upper Level convergence

Table 1. Selected Variables – CMB stands for Center of Maximum Buoyancy.

Due to the selection of variables that are highly uncorrelated with one another, each parameter selected identifies its own aspect of a storm. This is why the only significant correlation is the +0.65 correlation between the 0-CMB helicity and surface streamwise vorticity. That is because vorticity is a part of the calculation of helicity. CIN represents the negative forcing in the lower region of a sounding that can prevent a storm from occurring, but also often delays a storm from occurring until later in the day (Colby, 1984), when solar heating can increase the amount of energy available to a storm. Shear represents the change in wind direction and speed with height, which is necessary for a storm to become supercellular in nature. Wind shear can also determine a storm's ability to produce as mesocyclone (Wicker, 1996). Lifted Index describes the amount of positive buoyancy a storm will enjoy in the mid levels of the atmosphere. The rising motion at the level of free convection identifies lower level forcing, which is often required to overcome the negative buoyancy near the top of the boundary layer in many convective atmospheres. Helicity and surface streamwise vorticity both identify, in a slightly different way, the storm's potential to rotate at lower levels. Helicity is more likely to describe some kind of mid level rotation as well. Finally, the upper level divergence is largely related to the upper level dynamics, which is often associated with severe weather, especially earlier in the season.

With these seven parameters, many of the necessary ingredients for the formation of a severe thunderstorm are factored into the model procedure. Since the relationships between these seven parameters and each type of severe storm are slightly different, the different types of severe weather will be seen. However, it may become necessary to include one more variable for one or more type of severe weather to identify the different atmospheres more robustly.

4. THE MODEL PROCEDURE

The model procedure combines these seven parameters in a Bayesian Framework to devise a conditional probability of tornado occurrence, damaging wind occurrence, and hail occurrence. The Bayesian framework combines the parameters by adjusting the probability of severe weather occurrence for the value of each parameter added, starting at the base probability of severe weather.

$$P(\Theta \mid Y) = \frac{P(\Theta) * P(Y \mid \Theta)}{P(Y)} \qquad (2)$$

Bayes' Theorem (Equation 2) is used to shift the probability of a severe weather either upward or downward depending on the conditions described by each variable. For each type of severe weather, Bayes' Theorem is applied seven times, once for each variable. For each variable, the probability of a severe event, denoted by theta in the formula is multiplied by the probability that a severe weather even occurred with the variable's value in that range (the numerator), and then divided by the overall probability that the variable occurs in the range (the denominator). Each time Bayes' Theorem is applied for a variable, the distributions that produce the denominators for the other variables are also shifted. This is how the correlations between variables are accounted for to avoid over representing certain factors that lead to severe weather.

Although the procedure uses the same seven variables for tornadoes, damaging winds, and large hail, the relationships derived are slightly different. This allows the model procedure to differentiate between environments favorable to tornadoes and environments favorable to damaging winds, and likewise for large hail. However, since all three types of severe weather occur with the same basic conditions, many of the regions shown to have a high probability of one type of severe weather will also show a high probability of the other two types.

5. DISCUSSION

The model pinpoints regions of high likelihood of severe weather similar to the SPC convective outlooks. The main differences appear to be a tendency to pinpoint a smaller region for extremely high probabilities, and, of course boundaries in the likelihood contour zones that present themselves much less smoothly. The

	CIN	Shear	LI	W/LFC	Helicity	Vort.	Convergence
CIN	1	+0.26	-0.04	+0.16	+0.17	+0.15	+0.05
Shear	+0.26	1	-0.31	+0.25	+0.12	+0.05	+0.03
LI	-0.04	-0.31	1	-0.05	+0.05	+0.02	-0.16
W at LFC	+0.16	+0.25	-0.05	1	+0.17	+0.07	-0.23
Helcity	+0.17	+0.12	+0.05	+0.17	1	+0.65	-0.08
Vort.	+0.15	+0.05	+0.02	+0.07	+0.65	1	-0.02
Convergence	+0.05	+0.03	-0.16	-0.23	-0.08	-0.02	1

Table 2. Correlations between predictor variables.

main concern in the end results will be whether or not the procedure will be able to forecast all severe storms, and not whether or not the procedure will produce many false alarms. This is also the main problem with the previously derived Supercell Index (Wilt, 1994).

Further analysis of the results, predictability, and skill score are forthcoming. Upon this analysis, adjustments, such as the previously mentioned addition of another variable in the prediction of one or two of the severe weather types, or alterations in the relationships between the parameters, may prove to be necessary. With Bayesian Analysis, a higher percentage of storms will be predicted with the addition of more variables.

This procedure will be available for use with any numerical weather model that provides sufficient resolution, both spatially and temporally, to resolve necessary conditions for severe weather. This could provide assistance in forecasting severe weather likelihood on time scales anywhere from several hours out to a week ahead of time. This predictive index could also be used with climate models to forecast future severe weather climates under multiple climate scenarios, as is planned for future work.

6. CONCLUSION

Although there are some details to hammer out, and possible major adjustments, it seems very likely that a computer model procedure that uses atmospheric conditions described by predictive indices will provide a useful tool in the longer range forecasting of severe weather. This will provide an objective quantitative assessment of the likelihood of a severe storm at a given location at a given time, which will be based on computer model output. The use of ensembles, and multiple model formulations will further enhance the usefulness of this model procedure as a forecasting tool for severe storm forecasting.

As with model output in general, output from the model procedure is intended to be a tool to aid human forecasters, not replace them. Just as forecasters at the national weather service offices will see model outputted QPF and determine how likely it is to actually occur, forecasters at NWS, SPC, and other emergency management facilities, will be able to look at this product and determine how realistic the scenario is. The output from this procedure can only be as accurate as the model that produces the forecasts of the atmospheric conditions that are entered into the procedure. As improvements are made in both model accuracy, and in the procedures that produce these probabilities from the model outputs, the forecasts are bound to improve. Hopefully, this procedure that produces the probability of severe storms based on model output can become more and more accurate and useful over time as these improvements are made.

REFERENCES

Colby, F. P. 1984: Convective Inhibition as a Predictor during AVE-SESAME II. *Monthly Weather Review*, **112**, 2239-2252

Coloquhoun, J. R., and Riley, P. A., 1996: Relationships between Tornado Intensity and Various Wind and Thermodynamic Variables. *Weather and Forecasting*, **11**, 360-371

Mesinger, F., and coauthors, 2002: NCEP Regional Reanalysis. *AMS Symposium on Observations, Data Assimilation, and Probabilistic Prediction*, J59-J63

Rasmussen, E. N., and Blanchard, D. O., 1998: A Baseline Climatology of Sounding-Derived Supercell and Tornado Forecast Parameters. *Weather and Forecasting*, **13**, 1148-1164

Thompson, R. T., Edwards, R., and Hart, J. A., 2003: Close Proximity Soundings within Supercell Environments Obtained from the Rapid Update Cycle. *Weather and Forecasting*, **18**, 1243-1261

Weisman, M. L., Klemp, J.B., 1982: The Dependence of Numerically Simulated Convective Storms on Vertical Wind Shear and Buoyancy. *Monthly Weather Review*, **110**, 504-520

Wicker, L. J., 1996: The Role of Near Surface Wind Shear on Low-Level Mesocyclone Generation and Tornadoes. *18th Conference on Severe Local Storms*, American Meteorological Society, 115-119

Wilt, B. 1994: Supercell Index: prediction of supercells and tornadoes with model output diagnostics. Master's Thesis, University of Wisconsin-Madison