P 71

Testing and Application of a Forecast Confidence Index in a Variety of Regions and with Two Different Models

L. Phillips, M. Hjelmfelt, D. Kliche and W. Capehart
Institute of Atmospheric Sciences, South Dakota School of Mines and Technology

and

G. Kierstead and P. Pick
Environment Canada

Abstract

Accurate prediction of weather conditions has the power to save money and lives. Meteorologists play an important role in this process by assessing when numerical forecasts have a higher chance of failing. A way to assess the confidence one could place in a forecast is needed in situations where meteorologist input is not available. To address this issue, a Confidence Index(CI) algorithm was developed that uses rules to identify problematic features in a forecast. When these features are present, CI will reflect a lower confidence as a result. Previous implementation revealed CI's usefulness with the Canadian GEM model in Eastern Canada, but to demonstrate applicability over other geographical regions, testing was needed. This was done by calculating CI's for GFS High-Resolution Model forecasts in different locations. Results have revealed that despite changing models and location, the CI performed in accordance with the conceptual model. No significant differences in performance were found when switching models, but the algorithm struggled more over the Southeast US Region, possibly due to rules tailored for Eastern Canada. Thus, new rules and modifications to current rules have been developed to better account for a variety of locations and weather phenomena.

1 INTRODUCTION

Today, numerical weather prediction is heavily used for determining what is to come. However, the models are not without flaws, and at times can perform poorly. Forecasters that monitor these models can pick up on patterns and scenarios in which the models are more likely to fail, so many planners of outdoor events, construction projects, or other weather-sensitive activities rely on forecasters for an accurate forecast to guarantee their event is a success. However, decision makers are now more than ever using forecasts taken straight from the models to avoid the extra cost of a personalized forecast (Snellman, 1977). This leaves them susceptible to a risk if the forecast fails. Hence, a simple, quick calculation is desired to provide an assessment of the confidence one can place in a forecast. The impetus for the development of CI was for use with WRF forecasts for incident meteorology.

1.1 *CI Concept*

The CI v0.1 was designed by G. Kierstead and P. Pick to address this problem. It is calculated by an algorithm that uses a set of rules to identify problematic features in a forecast. The presence of more problematic features is reflected by a higher CI-Risk score or lower CI, hence a greater chance of the forecast failing. Ideally, the Confidence Index verses Model Error plot will be similar to Fig. 1. As pictured, high CI or low risk correspond with low model error. It is important to note that this relationship is not strictly linear. When the confidence index is low, it cannot be assumed that the model error will be high. The model may produce a good forecast for a challenging situation, but the chances of the model error being high are much greater.

* *Corresponding author address:* Lisa Phillips, Institute of Atmospheric Science, SD School of Mines, Rapid City, SD 57701-3995; e-mail: lisa.phillips@mines.sdsmt.edu.

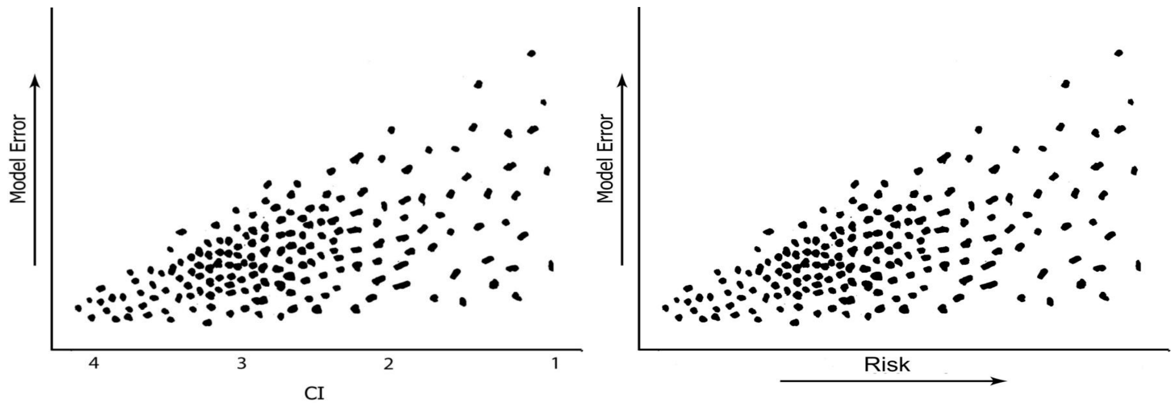

Figure 1: Left: Idealized results of Model Error vs. CI. Right: Idealized results of Model Error vs. CI-Risk. Low CI-Risk or high CI forecasts correspond to low model error, high risk or low CI forecasts correspond to a large variation of model error.

In examining the results in Fig. 1, it is apparent that a straight line could be fit to this data. A high correlation value is not expected since the data spread increases with increasing CI-Risk.

1.2 *Background*

Although the Confidence Index was not developed until recently, the idea that errors in forecasts tend to correlate with certain meteorological features has been used extensively in the forecasting community for a long time. Researchers often use a few different methods, such as the spread between ensemble members for a predictor of model skill (Hoffman and Kalnay, 1983; Leith, 1974). However, the Confidence Index does not utilize any predictors derived from ensembles since the available forecast may not have come from an ensemble. Since models may handle certain weather scenarios poorly, there is a fluctuation in model skill over time. From day to day, the same model can have very different skill scores due to the changing of weather patterns. Analysis of fluctuations in model skill can provide insight to the cause of these variations. If the causes for variations are related to meteorological flow or features, an assessment of the confidence can be found simply based on present features (Branstator, 1986).

1.3 *Research Objective*

This Confidence Index has the potential to be a useful tool for decision makers and forecasters alike. It should be especially useful with WRF forecasts for incident meteorology. Additionally, by examining the CI values for each forecast, decisions can be made regarding how many ensembles to run and how much time should be spent on a forecast. The usefulness of the CI hinges on the ability of this Confidence Index to perform as expected regardless of geographical location and model used to create the forecast. Hence, this research seeks to answer the following questions.

1. Can CI version 0.1 produce good output when used with other models?

2. Can CI version 0.1 produce expected results when used over different regions?

3. Can CI version 0.1 be improved by applying meteorological knowledge?

The purpose of this study is to assess whether or not the Confidence Index concept has potential beyond assessing forecasts from the GEM over Eastern Canada. If so, work to improve the CI will continue by examining the rules involved. Since the CI was developed for use over a specific area, certain meteorological phenomena may not be properly accounted for. Hence, new rule development will be important. Any possible improvements to rules will need to be tested and new rules will be created to account for more climatological regions.

2 PREVIOUS TESTING

2.1 *CI Calculations*

Kierstead and Pick performed preliminary tests to ensure that when used with real data, a distribution similar to Fig. 1 is produced. The original conceptual version of CI included 15 rules, but preliminary tests used only 5 of these. Problematic features or predictors described by these rules in CI v0.1 included closed low pressure centers, strong and very weak pressure gradients, large wind direction changes, and inconsistency between forecasts. Each rule affects both the CI and CI-Risk score. Since the CI value requires further efforts to calculate and the two values are closely related, CI-Risk will be referred to henceforth. The Closed Low rule works by examining the pressure field for areas of lower pressure surrounded entirely by higher pressure. If the lower pressure meets the criteria, it is considered a closed low and will contribute to a higher CI-Risk. The wind shift rule was designed to highlight areas of 500 hPa troughs/ fronts because these lead to a degraded forecast if mishandled. The rule should also have the capability to show the general location of the maximum wind shift. By examining the wind directions and comparing them to neighboring values, large differences in direction will cause the rule to penalize the forecast.

The Pressure Gradient rule handles potential forecast errors that tend to occur near gradients. When the location of a strong gradient is incorrect, large errors will be found in the forecast, especially for winds. Similarly, models struggle in cases with very weak gradients. This rule was designed to tag strong and very weak gradients and their presence leads to a higher CI-Risk. Inconsistency is highlighted in a forecast by the consistency rule. This rule compares the 12-hr and 24-hr forecasts that are valid for the same time. A model that does not have consistent solutions between model runs tends to be less accurate. Using these rules, Kierstead and Pick ran the CI v0.1 algorithm on approximately one year of Canadian High-Resolution Global Model forecasts over a 2000 x 2000 km region in Eastern Canada.

2.1.1 Verification Calculations

With the CI-Risk scores calculated, a measure of model error was needed. Since observations cannot be ingested into CI at this stage, a less than ideal method was employed. This was done by comparing the 500 hPa forecast pattern from the forecast to the same field for the model initialization, both valid at the same time. Although this method is not the most favorable, it does provide simple, quick assessment of model error. In the future, the verification process will be rebuilt to include multiple soundings at multiple levels.

2.2 *Statistical Multiple Regression Analysis*

Each problematic feature, or predictor contributes to the overall CI-Risk score, but some predictors are more influential than others. By conducting a statistical multiple regression analysis, different weights for each predictor can be established. The analysis was performed with the aid of Minitab. A regression equation was found by using a combination of variables from the CI rules that represent the presence and/or strength of a minimum gradient, maximum gradient, closed low, wind direction variables, and consistency.

By equating the response to model error, the variables from the rules are the predictors. In order to achieve an accurate way of relating model skill to CI score, the regression equation weights each variable differently. The P score indicates whether or not the variable has a significant correlation with model error. The lower the P score, the higher the correlation, and P-values below 0.025 are considered statistically significant. A variety of different combinations of the predictors were used in the regression analysis. The combination that was most statistically significant was considered to be the best representation of the results. The closer the *r*-value is to one, the better these predictors relate to the forecast error. Then the data were plotted with model error versus the outcome of the regression equation. A regression line is fit to the plot of the data (presented in this paper as a blue line).

From this we can examine the spread of the data and visualize possible problems with the CI algorithm and its rules. As a result of the spreading of data with increasing CI-Risk (decreasing CI), when fitting the data to a regression line, a high *r*-value is not expected. Regression analysis can aid in analysis, but does not describe the shape of the distribution. Thus, a line of fixed slope (presented here as a thick black line) will show the upper bounds of where we expect the data to fall. Between this line and the *x*-axis is considered the envelope. Points that fall outside of this envelope indicate the Confidence Index incorrectly diagnosed the forecast and did not penalize it sufficiently.

Further analysis involving the residuals may be implemented to gain more concrete results. Since the residuals represent the distance between a data point and the regression line and also account for the direction a data point is from the line (above or below), data points falling above the regression line by a certain distance or higher can be excluded. An average can be calculated to quantify this.

Results in Fig. 2 show good correlation between the CI value and the forecast error, providing a proof of concept for continuation of this work.

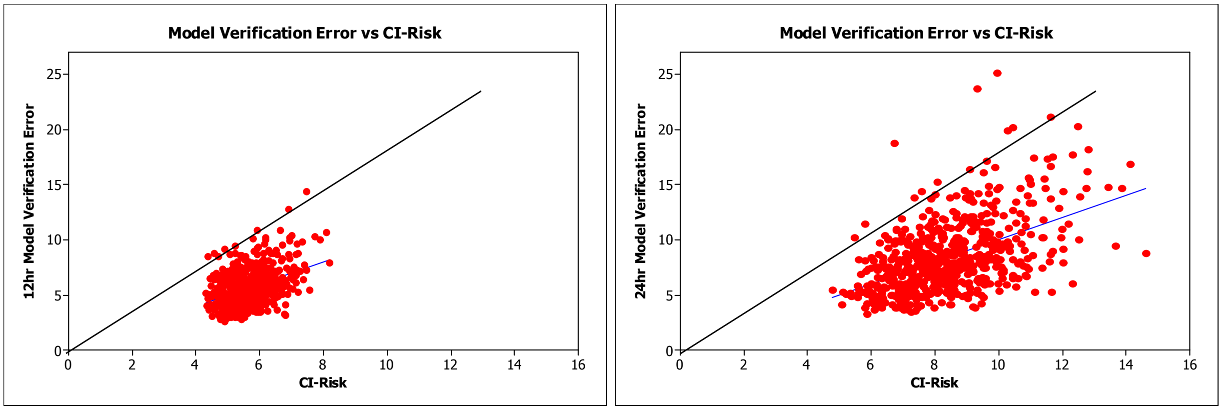

Figure 2: Results from preliminary testing using the East Canada region with the Canadian model. Left: 12-hr forecasts, Right: 24-hr forecasts. The black line represents the top of the envelope; the blue line is the regression line.

The trend of increasing maximum error with increasing risk is visible in this data and the distribution seems consistent with the idealized case. As shown, there is a large difference between the 12-hr forecasts and 24-hr forecasts. This shows that 12 hours is not sufficient time for the forecast to go astray. Also notice the envelope line which depicts the allowable maximum error. Only three forecasts fall far above this envelope for the 24-hr forecasts, and these represent forecasts that the algorithm failed to assign low enough CI-Risk scores.

3 PROCEDURE: VERSION 0.1 TESTED ON DIFFERENT REGIONS

3.1 *Data*

The CI algorithm runs on model data output at the 500 hPa level, and is created to handle data from a variety of current models. The second test of version 0.1 was run on the Global Forecast System high resolution 12- and 24-hr forecasts dating from 2 May 2008 to 30 April 2009. The GFS was chosen over the WRF model because it is already used operationally by our customer and it provides a worldwide data set that allows testing for geographic locations beyond North America. A few days are missing from the dataset, but not enough to be statistically significant. Using a year's worth of data will allow more accurate comparisons between this study and the previous.

3.2 *Regions*

In order to test the performance of version 1 of CI, the Confidence Index algorithm was run over multiple new geographical regions. For comparison purposes, a region over eastern Canada was created that corresponds to the area used in the previous study. Figure 3 shows the region used in the previous study and the region used in this study.

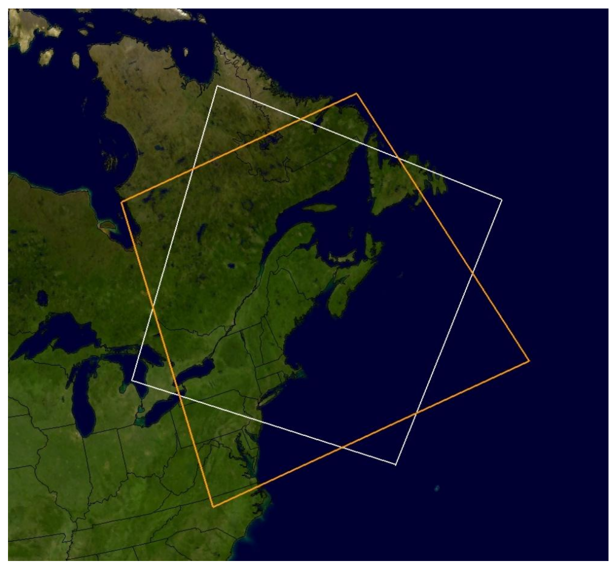

Fig. 3: Previous study area in white, present study area outlined in orange.

These regions could not be matched up perfectly due to differing map projections, the Canadian Model using a polar stereographic grid versus the GFS model's latitude-longitude grid. Four of the new areas are located in the United States and were chosen based on a familiarity of model error tendencies associated with the effects of specific topography and oceans. Multiple regions depicted in Fig. 4 were chosen to gauge how well CI handles different terrain, tropical systems, and data voids over the ocean. The Central region was created to examine relatively flat terrain with good data availability and lack of oceanic influence. Two different areas were examined out west, the Rockies and the West Coast. The main reason for including the Rockies region is to insure CI works over complex elevated terrain. The region was centered to minimize overlap with the Central region. The West Coast purposely includes ocean over the upwind portion of the domain and reaches into the Baja Peninsula for tropical storms as well as the highly active Pacific Northwest. To obtain an even better idea of how CI handles tropical systems and the data void of the ocean, an area in the southeastern U.S. (Southeast) was also chosen.

We also tested CI in the South-Central Asia region where we hope the CI will be used. This area will show the effects of complex terrain, oceans, and poor data coverage.

4 RESULTS

The same analysis method used for the preliminary test over eastern Canada was used for the multi-regional results. In examining the plots of results, it is important to note the number of forecasts that fall significantly outside of the envelope, the r- value, and the overall distribution of forecasts.

4.1 *Regional Analysis*

4.1.1 Eastern Canada

Results of this study from the region over eastern Canada are shown in Fig. 5. Two major differences exist between the previous test and the current results. Different models were used as well as different years.

Upon visualizing the plots, the lack of spread in the 12-hr forecasts is apparent. This is a repeating occurrence, and is thought to be because the forecast had less time to degrade; forecasts are generally good and receive low risk scores. It is likely that the spread is mainly due to noise. Little difference exists between the 12-hr forecasts from the Canadian Model and GFS Model. The 24-hr forecasts definitely display greater differences and for this reason will receive more attention.

The 12- and 24-hr Canadian and GFS model forecasts performed as expected, with a positive correlation between model error and CI-risk. A few outliers exist in the data, indicated by forecasts falling above the black envelope line. The Canadian 24-hr forecasts appear to have the greatest number of outliers, and contain three forecasts that fall well above the acceptable range. The general spread of the data is greater with the Canadian model as well-indicated by CI-Risk scores ranging from about 5 to 15. With the exception of outliers, it appears to fit the ideal results the best. However, when accounting for these outliers it

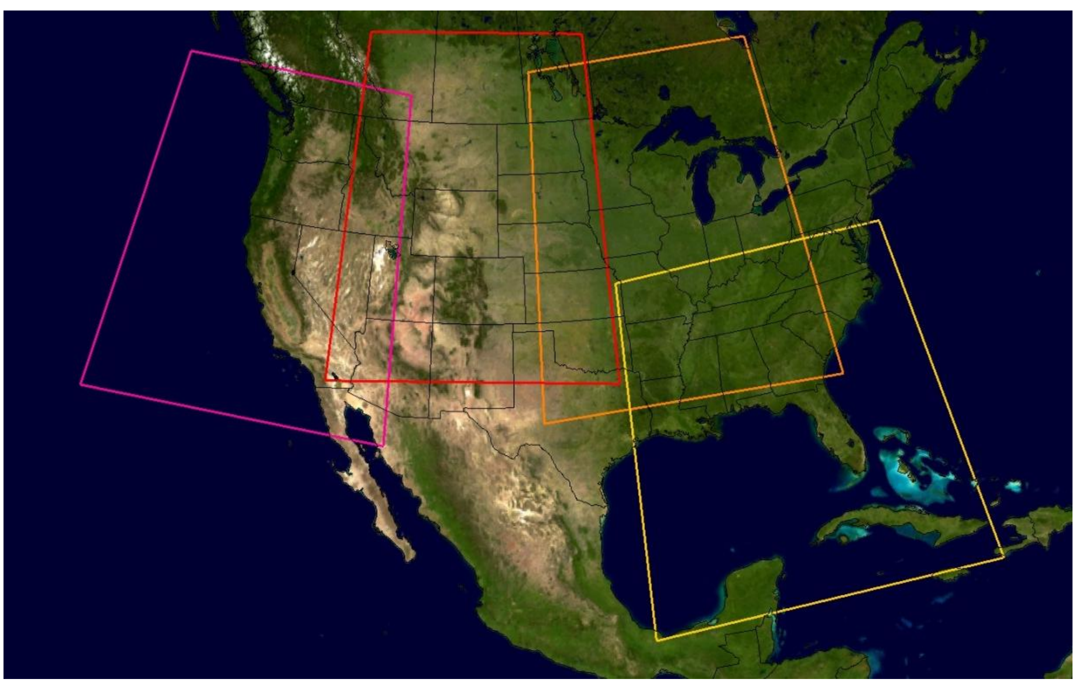

Figure 4: Left to right, West Coast, Rockies, Central, and Southeast regions.

seems the GFS 24-hr forecast fits closest to the ideal case. Examining the *r*-values can also show the spread of the data. High *r*-values indicate that the points tend to fall close to the regression line. As noted before, the ideal case is not expected to have a high *r*-value. Nonetheless, these *r*-values can help determine if forecasts are more clumped or spread out along and close to a line. In this case, the highest *r*-value 0.50 corresponds to the Canadian 24-hr forecast. It is interesting to note that for the GFS forecasts, the 12-hr forecasts scored higher than the 24-hr forecasts. However, this could be due to the fact that small CI-Risk values correspond with less variation of error.

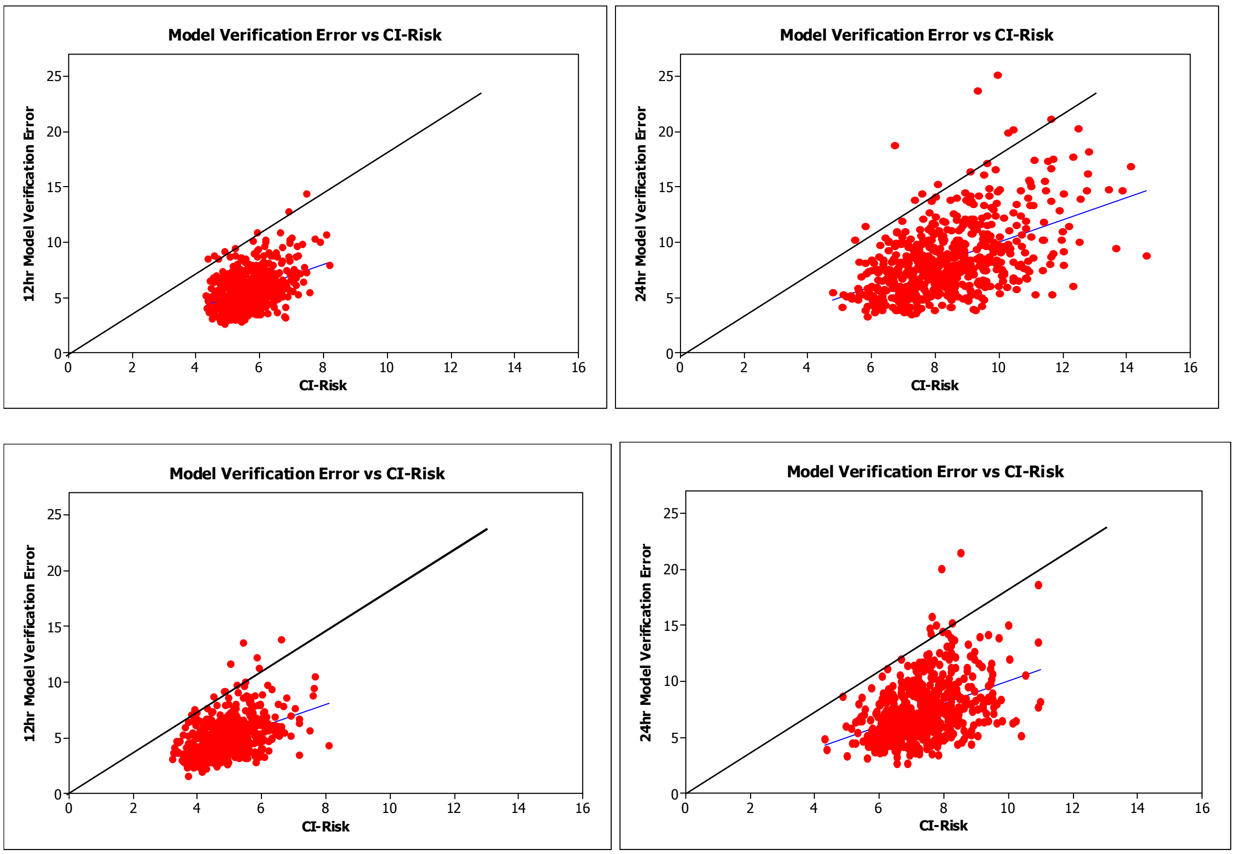

Figure 5: Scatterplots for forecasts over Eastern Canada. Top: Canadian Model results: left, 12-hr forecasts r = 0.41; right, 24-hr forecasts r=0.50. Bottom: GFS Model results: left, 12-hr forecasts r = 0.46, right, 24-hr forecasts r = 0.42. The black line represents the top of the envelope; the blue line is the regression line.

4.1.2 US West Coast

Summarized results for the West Coast region are found in Fig. 6. From the scatterplot, the expected trend is present with low model error corresponding to low CI Risk, and high CI Risk indicating more variable error. Of all the regions tested, the results for the West Coast have the greatest variation of CI Risk scores for both the 12- and 24-hour forecasts. In other words, more forecasts received high CI risk scores, and the majority of these forecasts fell below the envelope line. The 24-hr scatterplot shows many more forecasts above the envelope line; however none grossly exceed the envelope.

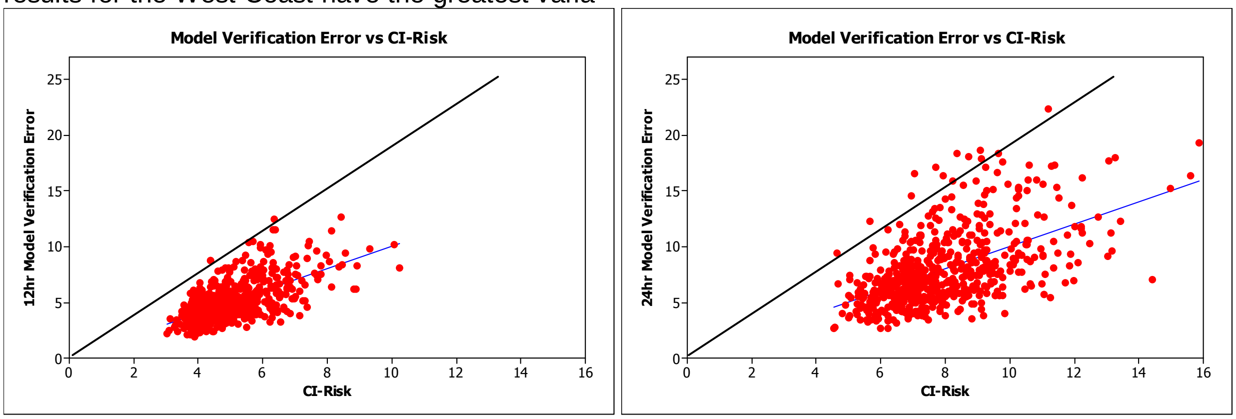

Figure 6: Scatterplots for the US West Coast. Left, 12-hr forecasts r=0.62; right, 24-hr forecasts r=0.55. The black line represents extent of the CI envelope and the regression line is blue.

4.1.3 Central US

Scatterplots of the results for the Central region are shown in Fig. 7. Error is low along with CI-Risk for the 12-hr forecasts, and forecasts tend to have little variation in CI-Risk scores. The distribution most closely resembles East Canada 12-hr scatterplots, and no significant outliers are present. The 24-hr forecasts have more spread and fewer forecasts fall above the envelope line. Only one forecast appears to be an outlier with a model error around 25. This forecast should have received a risk score of about 13 or higher, and hopefully new rules to be implemented will help correct this problem. *R*-values for the scatterplots seem to suggest that 12-hr forecasts are clumped close to a single point rather than scattered along and near the regression line. However, the differences between the *r*-values are rather small and may not be significant.

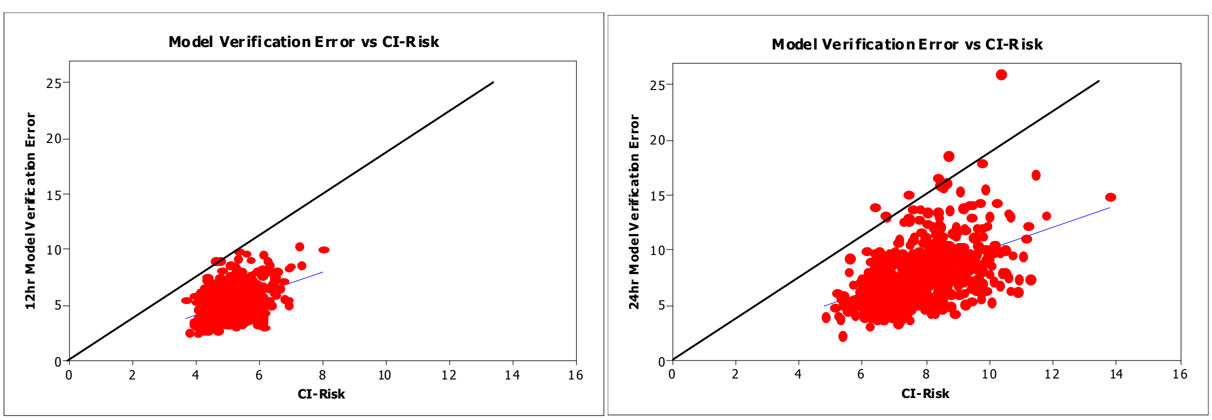

Figure 7: Scatterplots over Central United States. Left, 12-hr forecasts r=0.44; right, 24-hr forecasts r=0.49. The black line represents extent of the CI envelope and the regression line is blue.

4.1.4 US Rockies

Results for the Rockies (Fig. 8) display the similar trend of the 12-hr forecasts showing significantly less spread than the 24-hr forecasts. Despite this, the 12-hr scatterplot still shows more variation in CI Risk than the Central region 12-hr forecasts. Both 12- and 24-hr scatterplots fit the ideal case well. A few points fall outside of the envelope, but they are still close to the envelope line. Examining the *r*-values between the 12- and 24-hr data have similar values and is thus inconclusive. However, when comparing *r*-values to different regions, the 24-hr *r*-value of 0.57 is the highest of all the North America regions.

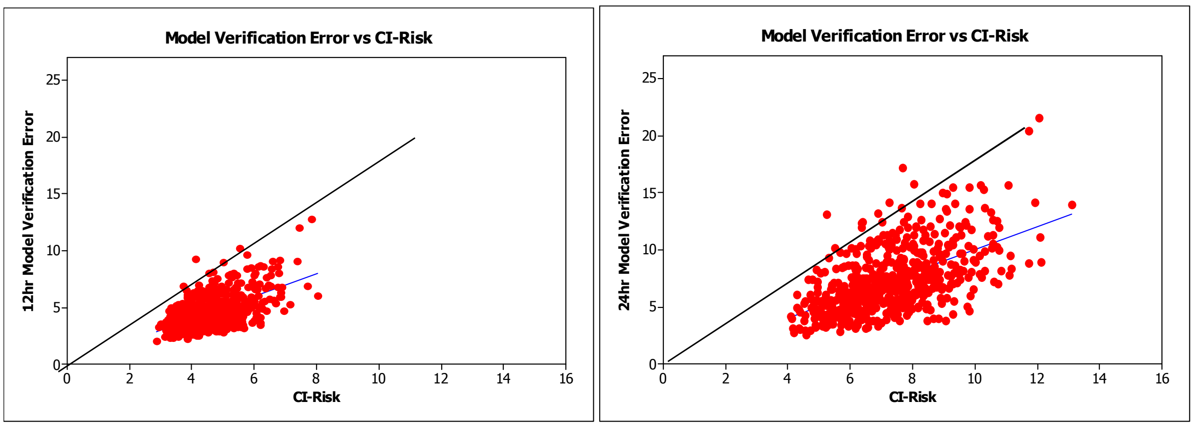

Figure 8: Scatterplots for the Rockies. Left, 12-hr forecasts r=0.60; right, 24-hr forecasts r=0.57. The black line represents extent of the CI envelope and regression line is blue.

4.1.5 Southeast US

Results from the Southeast region in Fig. 9 show slightly different results than other regions. The trend of having 12-hr forecasts with low Risk scores and low error remains the same, but in terms of the distribution fitting the ideal model they differ greatly. The 12-hr forecast seems to fit the ideal model well and closely resembles the West Coast 12-hr scatterplot. On the contrary, the 24-hr scatterplot shows a poorer fit to the ideal model. There is a smaller spread in CI Risk with the majority of forecasts earning a CI-Risk score above 6. Additionally, the forecasts that fall above the envelope line are clustered and tend to lie farther away from the line. It is possible that CI struggles more because of the lower latitude or because this region is more meteorologically active. Furthermore, no rule is in place to account for forecast errors caused by convection. By examining these forecasts that fall above the envelope, hopefully an origin for this error will surface and a new rule will be created for it.

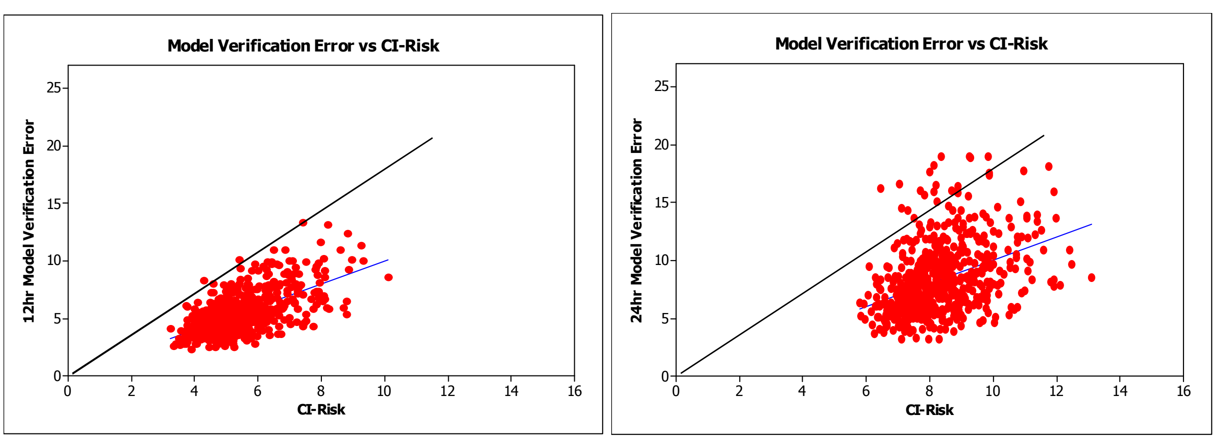

Figure 9: Scatterplots for the Southeastern U.S. Left, 12-hr forecasts r=0.63; right, 24-hr forecasts r=0.41. The black line represents extent of the CI envelope and the regression line is blue.

4.1.6 South-Central Asia

Results for the South-Central Asia region are also rather different than previous regions examined. In Fig.10, one can see that forecasts in this area tend to be good with low error. Features that the CI is programmed to find are not strongly present in these forecasts, as indicated by low CI-Risk scores. All points fall below this envelope. Despite the apparent differences between these and forecasts from other regions, correlations are similar to that of other regions. Some of this small spread may be caused because the lack of data in the region failed to pick up on errors occurring in data voids so errors were missed. Also, it appears this area is not very meteorologically active.

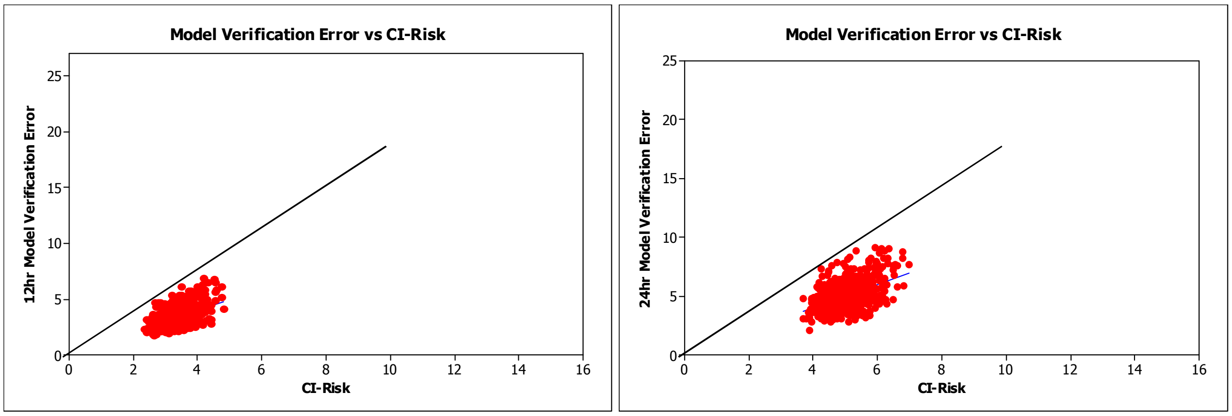

Figure 10: Scatterplots for South-Central Asia. Left, 12-hr forecasts r=0.52; right, 24-hr forecasts r=0.48. The black line represents extent of the CI envelope and the regression line is blue (mostly hidden beneath points).

4.2 *Equations and Coefficients*

In addition to using the linear regression analysis to determine *r*-values, portions of the regression equation can be combined for each region to create an average regression equation that is universal. Using the predictors and model verification error, a regression equation of the form below is created where FIELD_DIFF-VER is the verification error, 'a's represent constants, and GRADIENT_MIN, GRADIENT_MAX , MAXCLOSEDLOW and CONSISTENCY represent predictors.

$$\text{FIELD\_DIFF-VER} = a_1 + a_2\,\text{GRADIENT\_MIN} + a_3\,\text{GRADIENT\_MAX} + a_4\,\text{MAXCLOSEDLOW} + a_5\,\text{CONSISTENCY}$$

By creating an average value for each coefficient, an average linear regression equation was created for both 12-hr (Table 1), and 24-hr (Table 2) forecasts. Coefficients are indicated as an 'a' in Table 1 and 'b' in Table 2.

Table 1: Results for 12hr_FIELD_DIFF-VER

Region	a_1 coefficient	a_2 GRAD_MIN	a_3 GRAD_MAX	a_4 MAXCLOSEDLOW 8-NEIGHBOUR	a_5 CONSISTENCY	*r*
SC Asia (GFS data)	3.2073	-0.13091	-0.0269	0.0079	0.49939	0.62
Central US (GFS data)	3.3979	-0.01017	0.013109	-0.05055	0.26951	0.44
East Canada (GFS data)	2.4458	-0.06208	0.019082	0.04408	0.27029	0.46
US Rockies (GFS data)	1.4569	0.09418	0.027454	0.0488	0.35459	0.60
S-E US (GFS data)	2.2914	-0.14345	0.003923	0.09048	0.44517	0.63
US West Coast (GFS data)	2.008	0.10205	0.023273	-0.04695	0.40729	0.62
East Canada (GRIB_RGEM)	4.35	-0.168	0.0419	0.168	0.114	0.41
Average Coef.	2.736757	-0.045483	0.014549	0.037394	0.337177	0.54

Table 2: Results for 24hr_FIELD_DIFF-VER

Region	b_1 coefficient	b_2 GRAD_MIN	b_3 GRAD_MAX	b_4 MAXCLOSEDLOW 8-NEIGHBOUR	b_5 CONSISTENCY	*r*
SC Asia (GFS data)	3.58	-0.103	-0.0598	0.0683	0.674	0.48
Central US (GFS data)	3.69	-0.126	0.0424	0.104	0.429	0.49
East Canada (GFS data)	5.31	-0.304	0.0221	-0.0345	0.344	0.42
US Rockies (GFS data)	3.27	-0.247	0.0586	0.118	0.465	0.57
S-E US (GFS data)	4.66	-0.217	0.0094	0.181	0.454	0.41
US West Coast (GFS data)	4.25	-0.132	0.0367	-0.118	0.619	0.55
East Canada (GRIB_RGEM)	3.54	-0.216	0.141	0.477	0.264	0.50
Average Coef.	4.043	-0.192	0.036	0.114	0.464	0.49

An overall average is found by using the average coefficients for both the 12- and 24-hr, yielding the following formula. These coefficients are fixed when performing the regression.

$$\text{FIELD\_DIFF-VER} = 3.390 - 0.119\ \text{GRADIENT\_MIN} + 0.025\ \text{GRADIENT\_MAX} + 0.076\ \text{MAXCLOSEDLOW} + 0.401\ \text{CONSISTENCY}$$

Using this specific equation, the CI algorithm was run again over each region to compare results for this fixed average equation versus varying equations for each region. These results are given in Table 3.

Table 3: Correlation Coefficients for each domain using the Equation with Fixed Coefficients from both the 12 and 24-hr Forecasts

Region	*r* ï from average - 12-hr	*r* ï from average - 24-hr
SC Asia (GFS data)	0.49	0.44
Central US (GFS data)	0.43	0.48
East Canada (GFS data)	0.46	0.42
US Rockies (GFS data)	0.59	0.56
S-E US (GFS data)	0.61	0.40
US West Coast (GFS data)	0.61	0.54
East Canada (GRIB_RGEM)	0.35	0.43

Similarly, the CI algorithm was run again over each region using the average coefficients based on the 24-hr results. The results are given in Table 4.

$$\text{FIELD\_DIFF-VER} = 4.043 - 0.192\ \text{GRADIENT\_MIN} + 0.036\ \text{GRADIENT\_MAX} + 0.114\ \text{MAXCLOSEDLOW} + 0.464\ \text{CONSISTENCY}$$

Table 4: Correlation Coefficients for each domain using the Equation with Fixed Coefficients from only the 24-hr Forecasts

Region	*r* ï from average - 12-hr	*r* ï from average - 24-hr
SC Asia (GFS data)	0.47	0.42
Central US(GFS data)	0.43	0.49
East Canada(GFS data)	0.46	0.42
US Rockies(GFS data)	0.59	0.56
S-E US(GFS data)	0.60	0.40
US West Coast(GFS data)	0.60	0.54
East Canada(GRIB_RGEM)	0.36	0.45

Correlation values and regression equations using this Fixed Coefficients method are similar to using results achieved using the multiple regression analysis. The coefficients are a little lower for the SC Asia and East Canada data sets, but this is not a concern for SC Asia due to the small spread in the data. East Canada likely has a lower coefficient because of model differences.

4.3 *Further Work: New Rules and Modifications*

In order to improve CI v0.1 algorithms and ensure it will be useful over many regions, the interworking of the five individual rules was examined. From a meteorological perspective, areas for potential improvement were discovered. In addition, less than ideal CI performance in the southeastern region of the U.S. points to the need for a convection rule. The following is a brief description of modifications to certain rules and the new rules created.

4.3.1 Closed Low

Another version of the previous closed low rule has been created in hopes of capturing short waves, tropical systems that typically have a weak pressure signature, and accounting for system strength and size. To undertake this, instead of using the pressure field to find low systems, the new version will examine the relative vorticity field. This algorithm is designed to find areas of vorticity that exceed a given threshold. Once it finds these areas, it will search the surrounding neighbors for values that exceed the threshold as well. As long as the algorithm discovers that many neighbors do exceed the threshold, it will continue to examine neighbors farther and farther away from the point of interest. This allows the algorithm to only count large areas of relative vorticity. The maximum value of relative vorticity that lies within an area will also be noted.

4.3.2 Wind Change

The purpose of the original rule is designed to highlight areas of troughs and should be able to determine if the wind shift is downwind or upwind of a central point. An analysis was performed to test whether or not the algorithm would choose the wind shift associated with a 500hPa trough. Results showed that the algorithm could not correctly identify the most significant wind shift, and therefore a new version of this rule was created. This new version uses wind vectors instead of wind direction when comparing the winds at one grid point to another. It also uses averaging over larger areas to resolve wind shifts over larger areas instead of small, negligible differences between only a few wind vectors.

4.3.3 Geostrophic Wind

Working along with the wind change rule to highlight troughs/surface fronts is the newly created geostrophic wind rule. However, the key feature of this rule is that maximum magnitudes of geostrophic wind tend to mirror upper level jets. This is especially useful since wind forecasts have a greater chance of failing in the vicinity of jets. The rule works by calculating the geostrophic wind field, then calculates the magnitude for each geostrophic wind vector. Similar to the method used in the Closed Low Rule, it picks out the maximum value, and also searches for values that exceed a specific threshold. After finding these significant values, all the values of surrounding neighbors are checked for threshold exceedance. In this way, the large areas of high geostrophic wind magnitudes are revealed.

4.3.4 Convection

In order to account for poor forecasts due to convection, a rule was created to measure the amount of convection based on the lifted index that should be present at the forecast valid time. It uses the temperature and moisture variable at 500hPa and the temperature, pressure, and moisture variable at the lowest level. From these, a difference in virtual temperature, æT, between the two levels is found. The æT values are summed over the entire area, and the number of grid points that have positive æT values will be used to assess the Forecast Risk of the situation due to convection. Since data from only a few levels can be used for this calculation, this rule cannot account for convective inhibition. Hopefully this does not result in an over prediction of convection, but future versions will want to include another level if possible. Using only two levels also means only surface based convection is analyzed. Furthermore, it is uncertain how this rule will handle tropical convection that is typically characterized by small convective available potential energy values.

5 SUMMARY AND CONCLUSIONS

Results were found to be consistent between different forecast periods, years, forecast models, and regions. When comparing the 12 to 24-hr forecasts, there is a significantly higher error and higher CI-risk present for the 24-hr forecasts. There are differences between model and year, but these tend to be small. It appears that the GFS-Hires forecasts did better than the Canadian model with less points falling above the envelope. With consistency between models, CI is expected to work well with the WRF model. When varying regions over North America, generally small differences were found. The West Coast and southeast regions displayed the highest errors indicated by a large number of points above the envelope line. Since these regions cover a good deal of water, the upstream data void is a probable cause of this error. For the southeast, that performed the worst of the two, convection could be to blame along with its proximity to the equator. This led to the development of the convection algorithm, but a low latitude rule may be needed as well. Despite having little data and dramatic topography, the model seemed to do well over the Asia study area. However, the lack of data could mean the errors that occurred were not captured. This envelope used in the analysis appears to be similar for each data set, and appears to represent the maximum likely error for a specified confidence index value. Upon performing the fixed average regression analysis, it was discovered that this method is not significantly worse than using an individualized multiple regression specified for forecast period, region and model.

These results reveal that the Confidence Index works properly over different regions and models. In addition, a CI formula can be defined through a universal method of regression using fixed coefficients. This should work properly regardless of forecast period, region, or model. This Confidence Index appears useful, but does require some additional work with more algorithms needed to account for areas outside of the mid-latitudes.

Acknowledgements

This work was supported through US Army Contract W15QKN-06-D-0006 0007. We also thank the Canadian Metrological Center for providing access to their Global Environmental Multiscale model output. Thanks to Connie Crandall for manuscript preparation.

References

Branstator, G., 1986: The Variability in Skill of 72-hour Global-Scale NMC Forecasts. *Monthly Weather Review*, **114,** 2628-2639.

Hoffman, R. N., and E. Kalnay, 1983: Lagged average forecasting, an alternative to Monte Carlo forecasting. *Tellus A*, **35A,** 100-118.

Leith, C. E., 1974: Theoretical Skill of Monte Carlo Forecasts. *Monthly Weather Review*, **102,** 409-418.

Snellman, L. W., 1977: Operational Forecasting Using Automated Guidance. *Bulletin of the American Meteorological Society*, **58,** 1036-1044.