7 POST-PROCESSING OF CANADIAN REGIONAL-SCALE NWP TO DEVELOP FIRST-GUESS FORECASTS OF THUNDERSTORM AND SEVERE WEATHER THREAT AREAS

Neil M. Taylor[1], William R. Burrows[2,1], David M. L. Sills[2]

[1]Applied Environmental Prediction Science, Environment Canada, Edmonton, AB
[2]Cloud Physics and Severe Weather Research Section, Environment Canada, Toronto, ON

1. INTRODUCTION

Within Environment Canada (EC) efforts are underway to develop a next-generation forecasting, nowcasting, and alerting approach that will become the official source of weather prediction data and forecast products. Further, this approach will include the primary interface for forecasters to interact with data from observations, numerical weather prediction (NWP), and other sources. The primary means of forecaster data modification will likely be through the use of Meteorological Objects (MOs; Sills 2009; Sills et al. 2012); point, line, or areal depictions of meteorological features, processes, and alerts (e.g., watches and warnings). To assist in forecaster management of data and production of forecasts/alerts, they will have the option to view and modify "First-Guess" MOs (FGMOs). These FGMOs can be based on the forecast in effect, post-processed model data, climatology, science modules, or other sources and will be designed to support forecaster decisions as they generate forecasts and alerts. One requirement already identified is to have FGMOs available for alerts associated with thunderstorms, both in terms of storm vs. no-storm and storm vs. severe storm decisions.

Currently, the primary source of short-term NWP guidance within EC is the Regional Deterministic Prediction System (RDPS; Tanguay et al. 2012; Vaillancourt et al. 2012) with horizontal grid-spacing of 10 km. RDPS data provide national coverage and are used directly in the production of EC forecasts, either through forecaster modification or as automated products. NWP data at this scale are suitable for characterization of the storm or severe weather environment whereas storm-scale details (e.g., local storm intensity, convective mode) require kilometer-scale NWP such as EC's High-Resolution DPS (HRDPS; with 2.5-km horizontal grid spacing) or other models.

Given that RDPS data are already fully integrated into forecast operations at EC, and that post-processing of RDPS data is computationally inexpensive, we ask the question, "Can a single regional-scale NWP model be used to provide thunderstorm and severe weather forecasts that are a useful starting point (first guess) for the human forecaster?".

Our question is addressed in two stages: first, we calibrate RDPS output using observed lightning data to arrive at a probabilistic forecast of thunderstorm occurrence. Second, we combine the calibrated forecast with conventional severe weather forecast parameters to produce forecasts of severe weather occurrence. Data and methods are described in section 2 and the calibrated forecasts are introduced in section 3. In section 4 we discuss calibrated forecast verification followed by development of the severe weather forecasts in section 5. We conclude with verification of the severe weather forecasts (section 6) and a discussion of results and future work.

2. DATA AND METHODS

In our calibration procedure we utilize hourly RDPS output from 1200 UTC model runs covering the forecast period T+1 h to T+24 h. The RDPS runs at 10-km horizontal grid spacing and uses Kain-Fritsch convective precipitation parameterization (Kain and Fritsch 1990). For our calculations we use the lowest 58 vertical levels of the model. The period under consideration is 1 May to 30 September corresponding to the typical peak thunderstorm season over most of Canada. Lightning data (of any type) from the Canadian Lightning Detection Network (Burrows and Kochtubajda 2010) are utilized over the same period. For the present study we consider a domain encompassing the Canadian Prairies.

The objective of the calibration process is to identify appropriate predictors to use in the final

[1] *Corresponding author address*: Neil M. Taylor, Applied Environmental Prediction Science, Environment Canada, 9250-49th Street Edmonton, AB, T6B 1K5.
E-mail: Neil.Taylor@ec.gc.ca

forecast along with thresholds to apply for selected forecast probabilities. Initial identification of predictors (see section 2.1) was done through calibration when the RDPS had 15-km horizontal grid spacing (prior to Oct. 2012) and the same predictors and calibration procedure have been used in the present study with 10-km RDPS grid spacing.

2.1 General Calibration Method

Initially, 50 model parameters characterizing stability, moisture, vertical motion, and vertical wind shear were considered for the calibration. At each grid point in our domain, and for each forecast hour over the calibration period, the parameter forecast value was compared against observed lightning within a 60 km x 60 km box. The size of the "search" box was defined as being ± 2x the horizontal grid spacing in the model which at 15-km spacing corresponded to a search area of 2400 km^2 (for the calibration in 2014 using the 10-km model, the 2x grid-spacing search area was 1600 km^2; a 40 km x 40 km box as illustrated in Fig. 1). If a lightning flash was observed in the period T+0 to T+59:59 minutes then the forecast value at the central grid point is associated with a "yes" event. If no lightning is observed then the forecast value is associated with a "no" event.

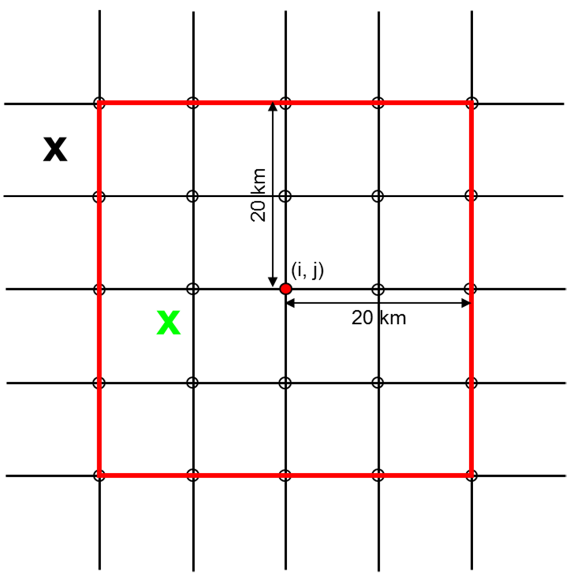

Fig. 1: Schematic representation of the search area (red box) relative to a model grid point using the 10-km RDPS. The forecast value is that of the central grid point (red circle). The green "X" represents a flash that would constitute a "yes" event. The black "X" is a flash outside the search box and would not be counted.

From the comparison against lightning, two datasets (yes and no) were compiled for each parameter and forecast hour. The datasets were compared first via boxplots to examine interquartile overlap, then by applying the Wilcoxon-Mann-Whitney (WMW) test in a manner similar to that used by Cohen et al. (2007) and following Wilks (2011). Converting the U-statistic from the WMW test to z-scores we selected a subset of parameters exhibiting both high z-scores and relevance with respect to a simple conceptual model for thunderstorm initiation (i.e., assessment of stability and vertical motion). Once identified, forecast thresholds for each parameter were derived from the decile values of the yes distribution for that parameter. This process is summarized graphically in Fig. 2.

The predictors identified from the above process are:

i) MUCAPE
ii) MUCIN
iii) MULPL-3 km CAPE
iv) Integrated vertical velocity below the MUEL (IVV)

Parameters i) and ii) are calculated in a conventional manner using the virtual temperature correction (Doswell and Rasmussen 1994). Parameter iii) is CAPE from the most-unstable lifted parcel level to 3-km above that level. For a surface-based parcel this is the same as the often used 0-3 km CAPE but the current formulation allows application to elevated parcels as well. Parameter iv) is the sum of upward vertical motion in the model column below the height of the equilibrium level for the most-unstable parcel. This parameter was developed to account for vertical motion contributions both within the boundary layer (e.g., associated with low-level convergence) and in the free atmosphere (e.g., associated with advection of positive vorticity aloft) and was found to highlight areas of ascent with spatial extent larger than those associated with low-level convergence boundaries alone.

As noted, the four predictors identified in discrimination analysis at 15-km were used for the calibration using the 10-km RDPS. For illustration, forecast thresholds derived from the 10-km calibration process for MUCAPE and MUCIN are shown in Fig. 3 for the T+1 to T+ 24 h forecast period.

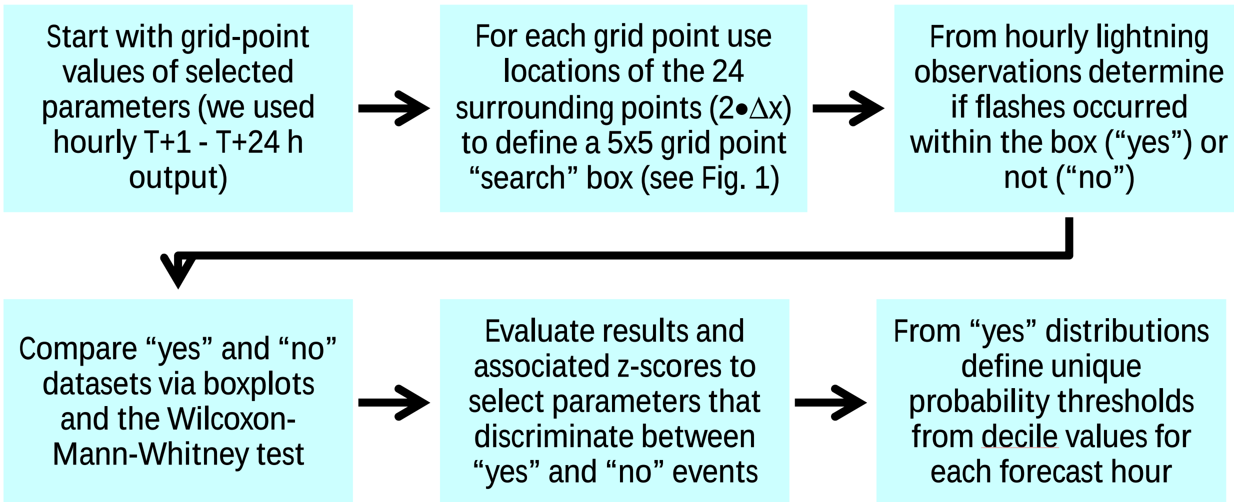

Fig. 2: Graphical summary of the calibration process used to identify suitable forecast parameters and associated thresholds.

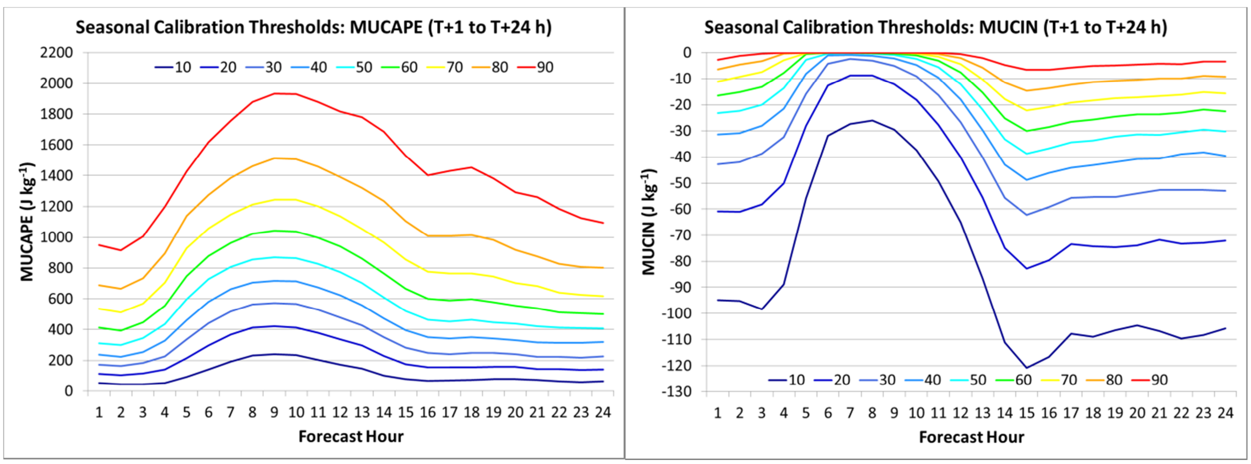

Fig. 3: Thresholds for MUCAPE (left) and MUCIN (right) for T+1 - T+24 hour forecasts based on seasonal calibration using 12 UTC runs of the 10-km RDPS from 1 May to 30 Sep 2013.

To obtain a positive forecast, we apply simple criteria whereby the thresholds for all four parameters must be met for a given forecast probability. While not an overly sophisticated approach, requiring thresholds to be met for all parameters at a given probability is an attempt to reduce the potential for false alarm forecasts.

Previous calibration tests (i.e., using the 15-km RDPS) were based on a full convective season (1 May to 30 Sep) of forecasts and lightning observations. This ensured that the intra-seasonal variability of parameter thresholds and model performance were captured. A limitation of this approach is that implementation of calibrated forecasts would be delayed following changes in NWP model characteristics. During the summer of 2014 we have tested an additional approach utilizing a "running" calibration based on thresholds and lightning observations from a prescribed number of days prior to the valid forecast time. The development of both types of forecasts is described below.

2.2 Seasonal Calibration

For seasonal calibration forecasts in 2014, NWP and lightning data from from 1 May to 30 Sep 2013 were used. In this case, forecast thresholds are derived from decile values of the yes-event distribution averaged over the entire season for each forecast hour (T+1 to T+24 h). As an example, parameter thresholds for the T+6 h forecast are derived from the average of T+6 h forecasts over the 153 specified days from 2013. For illustration, a sample forecast from the seasonal calibration is shown in Fig. 4 where forecast probabilities are colour-filled and lightning observations from the corresponding hour are overlaid.

2.3 Running Calibration

These forecasts are based on calibration periods of the previous 5, 10, 20, and 30 days in 2014 with thresholds recalculated daily. The use of multiple calibration periods is to test the performance of each forecast relative to the seasonal forecast and examine the number of days required to obtain a sufficient number of yes events for a useful forecast. During subjective evaluation over summer 2014 the general areal coverage of each of the running calibration forecasts was similar while the forecast probabilities within positive forecast areas varied between them. In some cases, especially early or late in the convective season, there were no yes events identified for a given forecast hour. In these cases, parameter thresholds from the seasonal calibration were substituted.

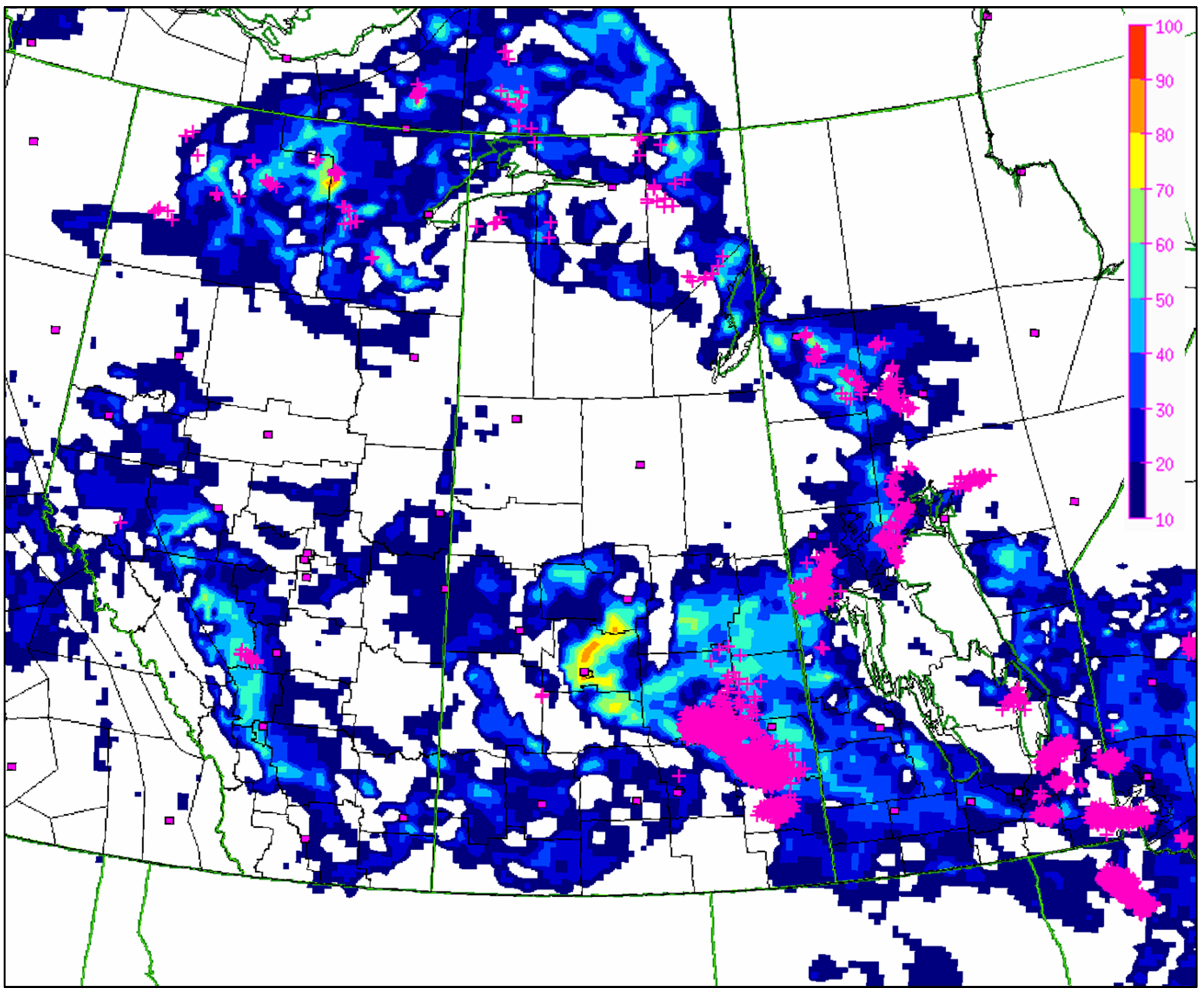

Fig. 4: Example seasonal calibration forecast based on the 10-km RDPS for T+12 h valid at 00 UTC 6 July 2014. Forecast probabilities range from 0-100 %. Lightning observed from 00:00:00 to 00:59:59 UTC is plotted as magenta crosses.

3. CALIBRATED FORECAST VERIFICATION

Verification for all calibrated forecasts is done on a 1-h basis for each forecast hour and each day. We compare the forecast for each forecast hour against observed lightning from T+0 to T+59:59 minutes. At each forecast grid point a search box of 60 km x 60 km is used to determine if a lightning flash was observed or not. This search area was used for consistency between verification at 10 km and 15 km. A positive forecast with at least one observed lightning flash is identified as a hit while a positive forecast with no observed lightning is a false alarm. Conversely, lightning observed with no positive forecast is a miss and cases with no lightning observed and no positive forecast are correct null events. This process is conducted for each forecast probability at a given grid point and a standard 2 x 2 contingency table (e.g., Wilks 2011) is populated based on the number of hits, misses, false alarms, and correct nulls detected. Since we are treating the calibrated forecasts in a probabilistic sense, we have selected Relative Operating Characteristic (ROC) curves and the Attributes Diagram to evaluate forecast performance (Fig. 5). In both diagrams in Fig. 5 we have included our previous calibration tests using the 15-km RDPS for comparison. In that test, we determined forecast thresholds based on the 153-day convective season in 2012 and verified the forecast using those thresholds for each day and forecast hour during the same period.

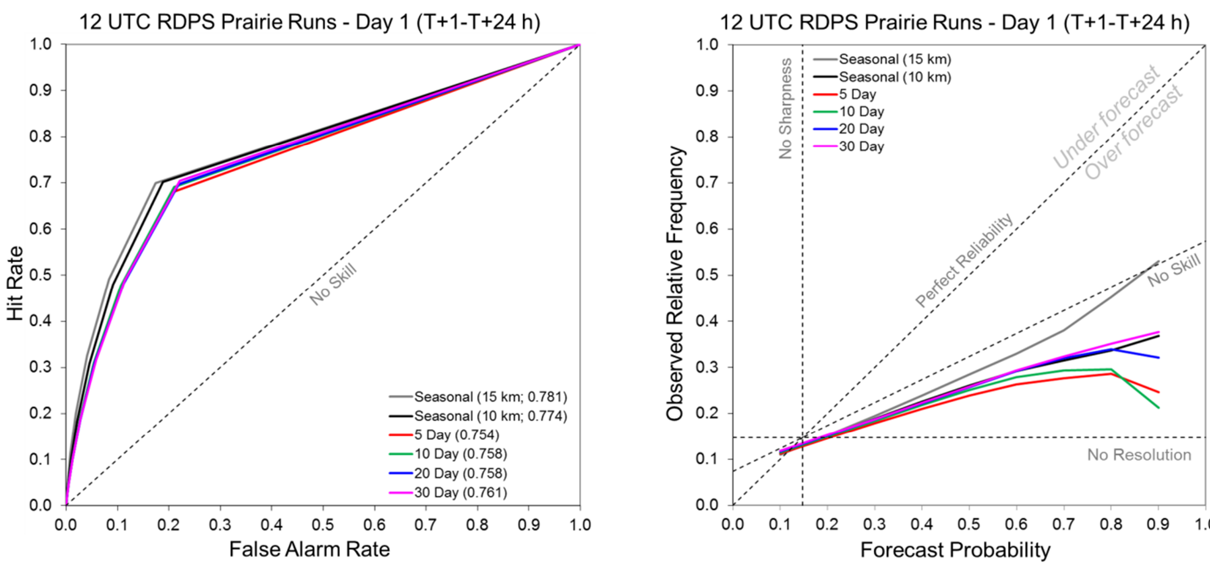

Fig. 5: ROC curves (left; ROC area in parentheses) and attributes diagram (right) for the seasonal (10 km [black] with 15 km [gray] for comparison) and running-calibration (10 km only) thunderstorm forecasts. No Sharpness and No Resolution lines in the attributes diagram are based on the sample climatology of the 10-km RDPS seasonal calibration.

Considering the ROC curves first we see that all forecasts exhibit some ability to discriminate between yes and no lightning events. ROC areas for all curves are ~0.8 with the seasonal calibration (both at 15 and 10 km) showing slightly better performance than the running calibration forecasts. Discrimination via the running calibration forecasts is nearly the same for all calibration periods with the 30-day period having the best ROC area overall.

The attributes diagram indicates that all the forecasts tend to over forecast thunderstorms. While this may be desirable in a general given the hazards associated with thunderstorm events, an improvement in overall skill may be required. The best reliability occurs for low forecast probabilities with reliability decreasing towards higher forecast probability. All the forecasts tend to exhibit resolution above sample climatology but without positive contributions to the Brier Skill Score (not shown). There is a general tendency for the forecasts at 10-km grid spacing to exhibit poorer reliability than at 15 km. We speculate that this difference may be largely accounted for by considering the search area boxes used for calibration in each case. For the 15-km model the 2x grid spacing results in a search area of 60 km x 60 km = 3600 km^2 while for the 10-km model the area is only 40 km x 40 km = 1600 km^2. The smaller area used for the 10-km model likely results in decreased hits and increased false alarm forecasts thus reducing the overall observed relative frequency.

A few additional comments on our overall verification approach are worth mentioning:

- Use of only 1-h forecasts and lightning observation periods are fairly stringent criteria with respect to the RDPS.

Relaxing these constraints will very likely improve overall verification scores.

- The verification considered all forecast hours collectively. Subjective evaluation of the RDPS and the calibrated forecasts have indicated a tendency for positive forecasts to occur much too soon in the forecast period relative to observed lightning. That is, conditions favourable for thunderstorm initiation occur in the RDPS before thunderstorms are observed; sometimes by several hours. This characteristic of RDPS performance is thought to significantly increase the number of false alarm forecasts. Since we have calibrated our forecasts to yes events only, this behavior is not accounted for in the calibration process.
- An increase in lightning search area for the calibration at 10-km is likely required.

Given a tendency for false alarms to occur early in the forecast period, and that more lightning is observed in the afternoon due to the diurnal cycle of storm activity, it may be reasonable to assume that forecast performance varies over the diurnal cycle. This is illustrated by contrasting the graphs in Fig. 5 against the attributes diagram generated for the T+9 h forecast only (Fig. 6; ROC curve was similar as in Fig. 5 and is not shown).

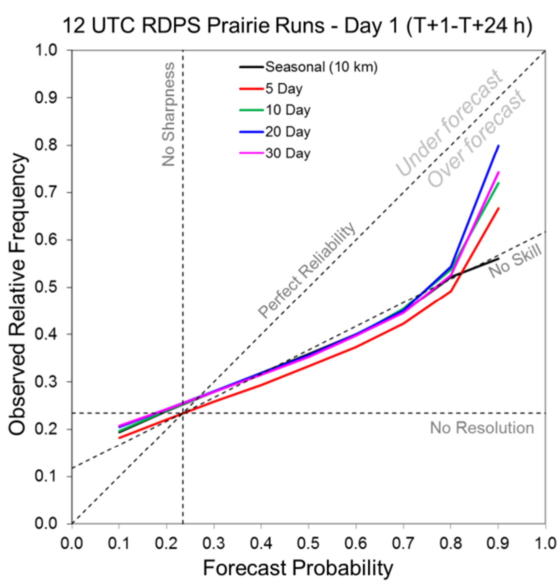

Fig. 6: Attributes diagram for all T+9 h forecasts (climatology defined by the T+9 h seasonal calibration data). Resolution and reliability are both improved when considering a single hour during the afternoon convection period over all the forecasts collectively.

4. SEVERE WEATHER THREAT AREAS

In support of severe weather watches and warnings it is desirable to provide for the forecaster a first-guess threat area. These areas could be modified by the forecaster and used to generate official watches as required. As a proof-of-concept experiment we have produced two products to support this. The first is a conditional first-guess severe weather threat area utilizing conventional parameters to help visualize the NWP characterization of a potential storm environment. The second is a severe weather occurrence forecast that combines the conditional area forecast with the ≥ 30% calibrated thunderstorm forecast used as a mask to highlight only regions where thunderstorms are likely. For tests discussed here, the thresholds used for the calibrated forecast came from the 2012 calibration of the RDPS when horizontal grid-spacing was 15 km.

The parameters and associated thresholds used for the conditional forecast are shown in Table 1. We define five categories based on convective mode or severe weather hazard and use colours to represent them in our forecast products. Thresholds used are based largely on those found in peer-reviewed literature and as recommended from the NWS Storm Prediction Center but with some adjustments to err on the side of detection and to accommodate thresholds sometimes used by forecasters on the Canadian Prairies. For general stability, MLCAPE is used unless the most-unstable lifted parcel level (MULPL) resides more than 500 m above ground in the RDPS. A positive forecast is generated when all thresholds are met for a given grid cell. Each successive category in Table 1 supersedes the one above it. An example of the conditional and occurrence forecast is shown in Fig. 7.

Table 1: Parameters and thresholds used to define severe thunderstorm categories in a conditional severe weather threat area forecast.

Category	Parameter	Threshold(s)
Severe Pulse / MCS	MLCAPE (MUCAPE MULPL ≥ 500 m)	≥ 1000 (1250) J kg^{-1}
	Effective BWD	< 30 kt
	Precipitable Water	≥ 20 mm
Convective Wind Gusts	DCAPE	≥ 500 J kg^{-1}
	Wind Index (WINDEX)	≥ 40 kt
Non-Supercell Tornado / Funnel Cloud	0-3 km MLCAPE	≥ 100 J kg^{-1}
	MLLCL	< 1500 m
	Effective BWD	< 30 kt
	Surface Relative Vorticity	≥ $8x10^{-5}$ s^{-1}
Supercell / Bow Echo	MLCAPE (MUCAPE if MULPL ≥ 500 m)	≥ 500 (750) J kg^{-1}
	0-6 km BWD (Effective BWD)	≥ 30 (30) kt
Supercell Tornado	MLCAPE	≥ 1250 J kg^{-1}
	MLLCL	< 1500 m
	Effective BWD	≥ 40 kt
	Effective SRH	≥ 150 m^2 s^{-2}
	0-1 km BWD	≥ 15 kt

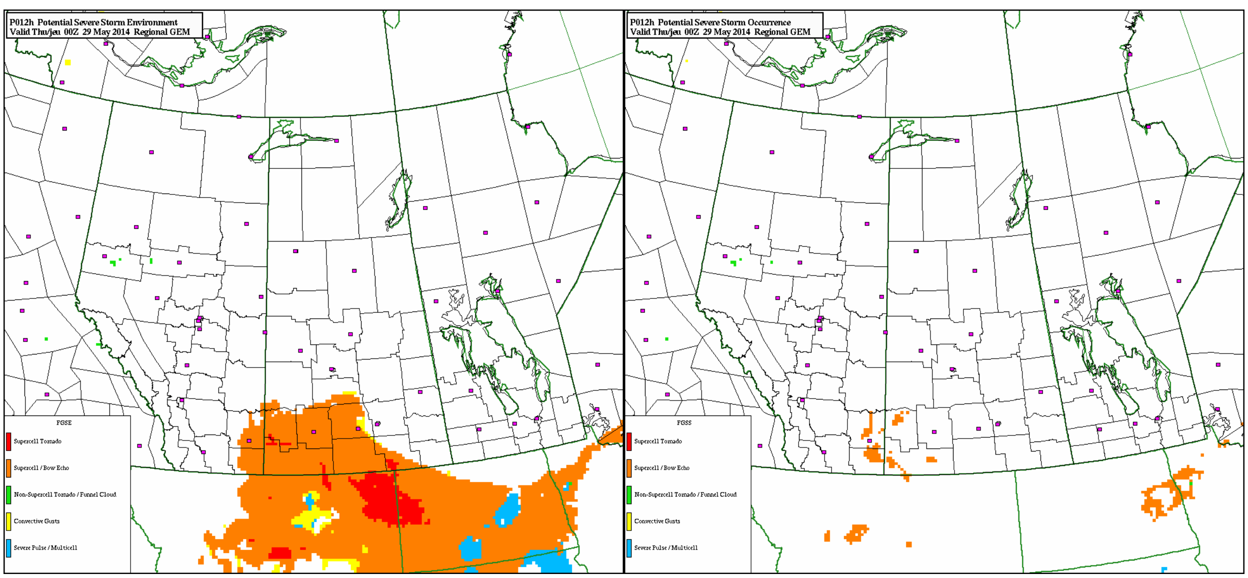

Fig. 7: First-guess severe thunderstorm areas by convective mode or weather hazard. Images are for T+12 h forecasts valid 00 UTC 29 May 2014. The conditional forecast (left) is produced using parameters and thresholds in Table 1. The occurrence forecast (right) masks the conditional forecast with the ≥ 30% calibrated thunderstorm forecast. Sub-regions within each province denote public forecast areas for which forecasts and watches are issued.

Both the conditional and occurrence forecasts were evaluated in real time by the authors during summer 2014 and were found to show utility in highlighting potential severe thunderstorm areas. To quantify performance, we have verified the occurrence forecasts against severe weather reports received by EC and severe weather watches issued by EC storm prediction centers on the Canadian Prairies.

Occurrence forecasts were verified for 45 days between 15 June and 31 July 2014 during the typical peak severe thunderstorm season over the Canadian Prairies. Only the "Day One" period was considered and defined as the period from 1200 UTC to 0600 UTC the following day (18 h). Over this period positive first-guess forecasts were identified if, within a public forecast region, there were at least eight contiguous grid cells defined by a single category for a period of at least two hours. This often resulted in more than one category being identified for a given region at some point over the 18 h period. For the verification scores discussed here we have removed forecasts of the non-supercell tornado / funnel cloud category and all reports of funnel clouds received by the SPCs. Similarly, we have not included any cases where funnel cloud advisories were issued by EC as these are not technically severe weather watches.

Verification was first conducted against severe weather reports. During the period in question 159 reports associated with 92 regions were received by the Prairie SPCs. Hits, misses, false alarms, and correct null forecasts were compiled and used to populate a 2 x 2 contingency table and generate several verification metrics (e.g., Wilks 2011). Since the first-guess forecasts are only valid when RDPS conditions are met, there is no lead time incorporated. We therefore allowed for a difference between report time and first-guess forecast valid periods of up to 1 h for a hit. Actual EC SPC watches were also verified against reports for the same period with no adjustment for report times vs valid time of watches (see Table 2).

Table 2: Verification scores for first-guess forecasts and Prairie SPC watches. Abbreviations for metrics are provided in the Appendix.

Verification	POD	FAR	HK	HSS	ETS	BIAS	ORSS	SEDI
First-guess vs Reports	0.58	0.95	0.34	0.05	0.03	12.7	0.62	0.47
Watches vs Reports	0.67	0.82	0.61	0.26	0.15	3.7	0.94	0.79
First-guess vs Watches	0.83	0.75	0.63	0.30	0.18	3.3	0.90	0.78

High FAR scores for both the first-guess forecasts and actual EC SPC watches are likely due, at least in part, to under reporting of severe weather events on the Canadian Prairies arising from low population density and limited infrastructure over large areas. The overall performance of the human forecaster is superior to the first-guess forecasts as expected. An additional contribution to first-guess performance may also be the apparent tendency for the RDPS environment to favor thunderstorms well before they occur (see section 3) leading to a substantial number of false alarms in situations that the human forecaster would not act on.

To gain a more complete appreciation of how the first-guess forecasts might perform, we compare them to watches issued by Prairie SPCs. We treat the watches as the observational dataset under the assumption that forecasters did not factor these products in their watch decisions.

This assumption seems reasonable since these products are only available to forecasters in an experimental capacity. In addition, the lead author worked full days in forecast operations for six weeks (including during the verification period) on the Research Support Desk (RSD; Sills and Taylor 2008) and did not observe these products to be part of forecaster's regular routines.

A simple approach was used whereby no temporal constraints were applied in comparison between first-guess forecasts and EC watches. If a forecast region was identified for a positive forecast using the first-guess product and a watch was issued at some point during the day-one period the forecast counted as a hit. This is again an attempt to account for lead time issues. At Canadian SPCs the issuance of watches does not necessarily depend on convective initiation. Often, watches may be issued before noon local time to coincide with regular public forecast updates. In these cases watches are intentionally issued several hours before convective initiation occurs. Similarly, watches may not be ended immediately after the severe weather threat is diminished for a variety of reasons. Comparing issue times of watches to first-guess severe forecasts is therefore unrepresentative of performance and in many cases produces an unrealistic number of missed events (not shown).

Scores from comparing first-guess forecasts to EC watches are shown in the bottom row of Table 2. Here we see fairly strong correspondence between first-guess forecasts and watches (e.g., POD = 0.83) though with some tendency to over forecast based on the BIAS score. A comparison to scores shown for first-guess forecasts vs. reports indicates much better performance against actual watches. False alarm performance is still less than ideal, partly for reasons already noted. Other metrics may be more appropriate for forecasting rare events such as the Symmetric Extremal Dependence Index (SEDI; Ferro and Stephenson 2011). In Table 2 we see SEDI scores for watches vs. reports and first-guess forecasts vs. watches both near 0.8. The most meaningful result here is that when first-guess forecasts are compared to watches, there is good overall agreement across several verification metrics. This indicates utility for the first-guess forecasts acting as an alert to forecasters of the potential for severe thunderstorms and possible watch requirements. Moreover, the first-guess forecasts may be suitable as a starting point for forecaster modification in the process of generating severe weather watches.

5. SUMMARY

A reasonably straightforward approach has been presented to generate calibrated forecasts of thunderstorm areas and to couple these forecasts with severe weather parameters to generate severe weather threat areas. The goal of this approach is to develop products that can be used as first-guess fields within the next-generation forecast approach under development within Environment Canada.

Four parameters (MUCAPE, MUCIN, MULPL-3 km CAPE, and IVV) were selected as predictors for the thunderstorm forecasts based on discrimination analysis, relationships to forecaster ADP, and a simple conceptual model for thunderstorm initiation. Calibrations were conducted using varying time periods of an entire convective season (153 days from 1 May to 30 Sep) and 5, 10, 20, and 30 days prior to the valid forecast time. All the calibrated forecasts were verified against observed lightning and indicate some skill in discriminating between yes and no lightning events (ROC areas) and resolution above sample climatology. A tendency has been identified for RDPS conditions to favour thunderstorm initiation before it occurs, frequently by several hours. This behavior is not addressed in our current calibration method and is responsible for a number of false alarms in the overall verification. Consideration of a single forecast time (e.g., T+9 h) illustrates that forecast reliability is significantly improved during the daytime convective period. Overall, the 30-day period shows slightly higher performance among all the running calibration forecasts evaluated.

Verification scores indicate little loss of performance in considering a 30-day calibration period against a full convective season. The 30-day calibration period allows for timely calibration updates in response to changes in the source NWP model. While the calibrated forecasts have shown utility in an operational setting, overall performance suggests that they may not be suitable to issue automated thunderstorm forecasts at this time. They may, however, be useful as a first-guess thunderstorm forecast to be modified by the human forecaster.

Conventional severe weather parameters and regionally-appropriate thresholds can be used to generate a conditional forecast of severe thunderstorm areas based on convective mode or severe weather hazard. When combined with the calibrated thunderstorm forecast, a severe weather occurrence forecast can be produced to highlight areas with the potential for thunderstorms and severe weather. Verification of the occurrence forecasts suggests that they compare well with actual watches issued by EC forecasters and may provide a suitable “heads-up” alert for forecasters or a first-guess forecast that can be modified to issue severe weather watches.

6. DISCUSSION AND FUTURE WORK

An attempt has been made to leverage the benefits of a regional-scale short-range NWP model for first-guess forecasts of thunderstorm and severe weather threat areas. The authors recognize that use of a deterministic convection-parameterized model may not yield the best objective results for thunderstorm forecasts. However, the benefits of taking this approach within EC are that:

- the RDPS is already integrated within operational forecaster workflow and software tools,
- horizontal grid-spacing is suitable for highlighting the storm environment without incorporating inconsistencies of deterministic forecasts of convective precipitation,
- post-processing of the RDPS is relatively computationally inexpensive,
- predictors for the calibrated forecast are easily relatable to observational data and the ADP process employed by the human forecaster, and

- subjective and objective verification results suggest there is utility in using this approach.

The results presented here should be considered preliminary. We are in the process of testing the current calibration approach with a larger search area for lightning and are also investigating the inclusion of lightning-specific predictors and a more sophisticated calibration process (e.g., classification and regression trees).

We have placed fairly stringent constraints on verification by only utilizing 1-hr forecasts and lightning observations. Experiments not discussed here have shown that relaxing the temporal constraints, e.g., by using 3-h maximum value forecasts and 3-h lightning observations, may yield better verification scores. One of the results sought from this study was to identify if a single running-calibration period could be identified for use and it was determined that hourly verification would be suitable for this.

An additional challenge to the current approach is the introduction of a national-scale version of the HRDPS at 2.5 km that will soon become the official source of deterministic NWP within EC. The authors expect that an alternate approach including gridding of lightning observations may be more suitable to ensure adequate correlation of lightning observations to forecasts at smaller horizontal grid spacing. Such an approach may also be applicable at 10 km.

Despite limitations associated with use of a deterministic regional-scale NWP model, the authors feel that we have demonstrated utility for our approach to yield first-guess forecasts of thunderstorms and severe weather areas. Further work will refine our methods and could form the basis of first-guess MetObjects for thunderstorms and severe weather within Environment Canada. Application to higher-resolution NWP models and ensemble prediction systems may also be explored.

Acknowledgements

This work has benefitted significantly from conceptual and technical discussions with our Environment Canada colleagues. Specifically, Julian Brimelow, Ron Frenette, Daniel Figueras, and Ron Goodson have provided valuable input and technical support. The authors wish to thank the management team within Environment Canada for continuing to support research and development in the areas of severe weather forecasting and Meteorological Objects.

APPENDIX

A brief description of the objective verification metrics used in this study is given in the table below. Descriptions are taken from, "Forecast Verification: Issues, Methods, and FAQ" by the WWRP/WGNE Joint Working Group on Forecast Verification Research. Available online at: http://www.cawcr.gov.au/projects/verification/.

Name of Metric	Acronym / Abbreviation	Brief Description
Probability of Detection	POD	Fraction of observed "yes" events that were correctly forecast
False Alarm Ratio	FAR	Fraction of the predicted "yes" events that did not actually occur
Hanssen and Kuipers Discriminant	HK	Skill in separating "yes" events from "no" events.
Heidke Skill Score	HSS	Accuracy of the forecast relative to random chance.
Equitable Threat Score	ETS	Correspondence of forecast "yes" events to observed "yes" events accounting for hits due to chance.
Bias	BIAS	Forecast frequency of "yes" events to observed frequency of "yes" events.
Odds Ratio Skill Score (Stephenson 2000)	ODDS	What was the improvement of the forecast over random chance?
Symmetric Extremal Dependence Index (Ferro and Stephenson 2011)	SEDI	What is the association between forecast and observed rare events?

REFERENCES

Burrows, W. R., and B. Kochtubajda, 2010: A decade of cloud-to-ground lightning in Canada: 1999-2008. Part 1: Flash density and occurrence. *Atmos.-Ocean*, **48**, 177-194.

Cohen, A. E., M. C. Coniglio, S. F. Corfidi, and S. J. Corfidi, 2007: Discrimination of mesoscale convective system environments using sounding observations. *Wea. Forecasting*, **22**, 1045-1062.

Ferro C.A.T., and D.B. Stephenson, 2011: Extremal Dependence Indices: improved verifiation measures for deterministic forecasts of rare binary events. *Wea. Forecasting*, **26**, 699-713.

Doswell, C.A. III, and E.N. Rasmussen, 1994: The effect of neglecting the virtual temperature correction on CAPE calculations. *Wea. Forecasting,* **9**, 619-623.

Kain, J. S., and J.M. Fritsch, 1990: A one-dimensional entraining/detraining plume model and its application in convective parameterization. *J. Atmos. Sci.*, **47**, 2784-2802.

Sills, D. M. L., and N. M. Taylor, 2008: The Research Support Desk (RSD) initiative at Environment Canada: linking severe weather researchers and forecasters in a real-time operational setting. Preprints, *24th Conference on Severe Local Storms*, Savannah, GA, Amer. Meteor. Soc., 9A.1. [Available online at: http://ams.confex.com/ams/pdfpapers/142033.pdf]

Sills, D. M. L., 2009: On the MSC Forecasters Forums and the Future Role of the Human Forecaster. *Bull. Amer. Meteorol. Soc.,* **90**, 619-627*.*

Sills, D., N. Driedger and W. Burrows, 2012: Verification of forecaster-generated iCAST thunderstorm nowcasts and comparison to automated thunderstorm forecasts: preliminary results. *Extended abstracts, 3rd World Weather Research Programme Symposium on Nowcasting and Very Short Range Forecasting,* Rio de Janeiro, Brazil, Paper 11.4. [Available online at: http://www.yorku.ca/pat/research/dsills/papers/WSN12/WSN12_iCAST_Verif_FINAL.pdf]

Stephenson, D.B., 2000: Use of the "odds ratio" for diagnosing forecast skill. *Wea. Forecasting*, **15**, 221-232.

Tanguay, M., L. Fillion, E. Lapalme, M. Lajoie, 2012: Four-Dimensional Variational Data Assimilation for the Canadian Regional Deterministic Prediction System. *Mon. Wea. Rev*., **140**, 1517–1538.

Taylor, N. M. and W. R. Burrows, 2010: Using Canadian GEM output for forecasts of thunderstorm initiation on the Canadian Prairies. Preprints, *25th Conference on Severe Local Storms*, Denver, CO, Amer. Meteor. Soc,. 3A.3. [Available online at: https://ams.confex.com/ams/pdfpapers/176105.pdf]

Vaillancourt, P. A., L. Fillion, J. St-James, A. Patoine, M. Reszka, M. Tanguay, E. Lapalme, B. He, M. Lajoie, A. Plante, A. Zadra, R. Mctaggart-Cowan, C. Creese, L. Lam, A. Rahill, S. Roy and V. Thomas, 2012: Improvements to the Regional Deterministic Prediction System (RDPS) from version 2.0.0 to version 3.0.0. Environment Canada Technical Note, 78 pp. [Available online at: http://collaboration.cmc.ec.gc.ca/cmc/CMOI/product_guide/docs/lib/op_systems/doc_opchanges/technote_rdps300_20121003_e.pdf]

Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences.* Academic Press, 676 pp.