6A.5 PROJECT PHOENIX - OPTIMIZING THE MACHINE-PERSON MIX IN HIGH-IMPACT WEATHER FORECASTING

Patrick J. McCarthy, Dave Ball, Willi Purcell
Prairie and Arctic Storm Prediction Centre
Meteorological Service of Canada

1. INTRODUCTION

In the numerical model-dominated weather services of the world, do meteorologists still add value in weather forecasting?

The forecast process within an operational weather centre is based upon the analysis, diagnosis and prognosis of skilled meteorologists. To aid their efforts, the forecasters use a wide range of tools including weather satellites, radar, workstations, and numerical weather prediction (NWP) models. Tremendous gains in NWP performance over the past 2 decades has made this tool typically the most utilized in the forecaster's toolbox.

As early as 1969 (Klein, 1969), it has been envisioned that computer-generated forecasts would increasingly dominate weather prediction. In 1970, NWP-generated text forecasts were being produced experimentally (Glahn, 1970) with the objective to reduce forecaster workload related to "routine" weather.

The seductive nature of the ever-improving NWP output raised concerns that forecasters were becoming too dependent upon this tool. Leonard W. Snellman noted this development as early as 1977 (Snellman, 1977). He coined the term "meteorological cancer" to describe the increasing tendency of forecasters to abdicate practicing meteorological science, while becoming just a conduit for information generated by computers.

By the late 1990s, managers at the Prairie Storm Prediction Centre (PSPC) were concerned that the performance of their forecasters was now only on par with the now prevalent NWP forecast products. An experiment, called Project Phoenix, was designed in 2000 to test whether forecasters still added value to the routine public forecast. The project in early 2001 and when unexpected results were achieved, the project was rerun later that year producing similar results.

This paper examines the results of Project Phoenix, and offers explanations for the results obtained.

2. PROJECT PHOENIX

Project Phoenix has two principal components: 1) operational forecasting, and 2) performance measurement and feedback.

OPERATIONAL FORECASTING

To test whether the PSPC meteorologists overuse NWP at the expense of their meteorological skills, or whether forecasters could no-longer add value, it was decided that a second "weather center" was needed. The project would compare the performance of the Phoenix team to that of the PSPC, plus an automated NWP-based text product.

The Phoenix weather center mirrored the PSPC. The PSPC was responsible for the forecasts, Watches and Warnings of the three Canadian Provinces. This area of responsibility represented one of the largest in the world, encompassing roughly 25% of the area of the 48 contiguous U.S. states. The PSPC forecast area was divided into 102 forecast regions. Three or four forecasters were normally on duty at the PSPC. The Project Phoenix teams operated their weather center "down the hall" from the PSPC with 3 forecasters and produced the same primary forecasts as the PSPC. The forecasters selected for the Phoenix team were chosen to best represent a mix of both experience and inexperience. Selecting a

Corresponding author address: Patrick McCarthy, Prairie and Arctic Storm prediction Centre, 123 Main Street – Suite 150, Winnipeg, Manitoba, Canada, R3C4W2; email: patrick.mccarthy@ec.gc.ca

team of “all-stars” was counter to the goals of the project and was intentionally avoided.

All versions of Phoenix followed the same general process, on a Monday-to-Friday basis. The forecast team worked from 8:00 a.m. until 5:00 p.m. and was responsible for issuing all of the regularly scheduled forecasts due near 11:00 a.m. and 4:30 p.m., in accordance with the official filing times for each product.

PSPC forecasters usually have a text product from the previous shift in place when they work through their routine. To maintain a similar approach between the two forecast teams, while avoiding cross-contamination between the two, the Phoenix forecasters began with the automated NWP-based text product produced earlier that morning near 00:00 A.M. This was the “SCRIBE” forecast. SCRIBE (Boulais, 1992, Boulais et al, 1992, Verret et al, 1993, Verret et al, 1995) takes direct model output from the Canadian operational model plus output from a number of NWP-derived statistical packages to generate a matrix of weather concepts. Text forecasts can be generated from these concepts. Forecasters can modify these concepts to generate new forecasts. An analogue to this system would be the IFPS (Ruth, 2002) and the NDFD (Glahn et al, 2003) in the U.S. National Weather Service.

Based upon their analysis, diagnosis, and prognosis efforts, the forecasters could alter the existing forecasts or start anew. Other than the raw SCRIBE forecast, the Phoenix forecasters had no access to any model or statistical information, while all official forecast products were also barred. This lack of model and human forecast data was the only significant difference between what was available to the Phoenix team and the PSPC forecasters. The Phoenix forecasters had to meet the same forecast deadlines as their PSPC counterparts. All non-transmitted forecasts valid at the deadlines for both forecast teams were deemed final and assessed. Forecasts for days 1 to 5 could be altered.

At the end of each day during the forecasting portion of Phoenix, the meteorologists offered a brief commentary on their shift, outlining any challenges that were faced and potential concerns with their forecasts.

The Phoenix, SCRIBE and PSPC forecasts were objectively assessed early the following morning using the Phoenix verification method. Feedback was immediately provided as raw numeric scores for each version, along with a written commentary outlining where each group had achieved superior performance, or had failed to match the accuracy of the other two forecast systems. Both the Project Phoenix and PSPC forecasters were aware of the details of the verification scheme used.

VERIFICATION

A detailed description of the Phoenix verification system is beyond the scope of this paper, though one is available (Ball, 2007). The following is a brief summary of the verification system.

A prerequisite to any comparison of forecasts is a verification system that will provide an accurate and meaningful measure of error (e.g. Stanski, 1982, Stanski et al, 1989), preferably in a concise fashion that the end user, in this case the general public, can readily understand. Many systems of verification have traditionally been used by Environment Canada, but most concentrate only upon one or two readily quantifiable parameters, such as a daytime high. These methods fail to provide the complete story and the utility of the information can be masked in statistical parlance.

Project Phoenix adopted a verification scheme that attempted to measure the severity of forecast error and its impact upon the residents of the Prairie Provinces, drawing heavily upon public responses to recent Environment Canada opinion surveys (e.g. Environics Research Group, 1999, Angus Reid Group Inc., 2000). In effect, the system was designed to recognize that missing a five centimetre snowfall in a large centre would result in a far greater negative reaction than would a forecast missing a few late-afternoon sunny breaks in a rural region.

Forecasts of precipitation, wind and sky conditions were assessed, based on a five-category, nonlinear error scale (Figure 1). Forecasting the correct category resulted in a zero-error score, while an error of four full categories resulted in an error charge of 100 percent. Due to its emphasis on improving forecasts in the shorter term, the Phoenix verification system assessed the forecasts in greater detail than has typically been the case, including the timing of significant events.

Forecast Category	Precipitation (Rain)	Precipitation (Snow)	Wind Speed	Sky Condition
1	Nil	Nil	10	Sunny
2	0.3	0.3	20	Sunny with cloudy periods
3	0.6	0.6	30	Mix of sun and cloud
4	>= 0.2 cm Showery	0.2 cm - 2 cm	40	Cloudy with sunny periods
5	>=0.2 cm Extensive	> 2 cm	50+	Cloudy

Figure 1. Verification Categories for precipitation, wind speed and sky condition.

Temperature was handled differently. The public surveys indicated that the Canadian public could tolerate temperature errors of four degrees Celsius or less but that errors in excess of 4 degrees were deemed unacceptable. Temperature forecasts were verified on a scale that was linear, but with a major discontinuity at five degrees Celsius of error. Temperature errors of four degrees or less incurred low to moderate error scores, while departures of five degrees or more suffered high error charges. Temperatures within one degree were deemed error free, while temperature errors of 10 or more degrees incurred an error assessment of 100 percent (Figure 2).

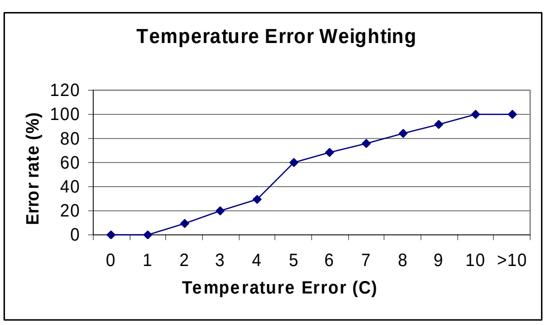

Figure 2. Temperature error weighting. The difference between the forecast maximum or minimum temperature and the observed temperature (rounded to the nearest whole number) is assigned an error rate.

The error scores were combined into one holistic error. Since the same public surveys noted that the various weather elements ranked differently in importance, each weather element was weighted differently. Based on public opinion, precipitation forecasts account for 45 percent of the forecast score and temperatures were ascribed a 25 percent proportion. Wind velocities contribute 15 percent of the forecast value, with sky condition and obstructions to vision accounting for the remaining 10 percent of the total score.

The verification system is heavily weighted toward routine weather events due to the limited duration of Project Phoenix and its primary objectives. Specifically, there was no attempt to differentiate between extreme events that would warrant a weather warning. A Phoenix verification system for extreme weather and warnings has since been created.

The system also used a weighting factor for each forecast region, based in part upon population, but also considering factors such as infrastructure, industry and geographical area. As a result, a major urban centre could have a weighting of up to five times that of a small rural region.

Each forecast part period was weighted differently. The public surveys indicated that most people made the majority of their final weather-affected decisions based upon the first part-period information, while decisions were much less likely for subsequent part-periods. Therefore, the “today” forecast was weighted much more heavily than the “tomorrow” forecast if the scores were to be combined. Both “combined” scores and individual “part-period” scores were provided to forecasters. Inevitably, forecasters devoted more attention to the first part-period of their forecasts. The information discussed here reflects the individual part-periods.

The scores were calculated at the predetermined verification sites for each of the Phoenix forecasts, the automated product generated by SCRIBE, as well as the official forecasts simultaneously issued by the PSPC, producing a single error percentage for each of Phoenix, SCRIBE and the PSPC.

All of the forecast parameters were then manually reviewed and assessed against actual observations, and where required, radar and satellite imagery. Results for individual sites, weather elements and times were kept, but the tallies were also combined into a single forecast error score that typically ranged from near zero for an excellent forecast, to in excess of 25 percent for what would typically be perceived as a "bust" forecast across a significant part of the region. Based upon the verification's categorical errors which were based upon public surveys, the majority of the public should begin to notice a forecast was in error to some degree around the 5 percent error mark, with increasing dissatisfaction evident for error scores above 10 percent.

3. VERIFICATION RESULTS

The lack of access to model guidance was the only significant difference between the forecasting experience of the Phoenix teams and the PSPC meteorologists. Both groups of meteorologists had the ability to adjust the numerical weather prediction guidance on the basis of recent real-time data, while the SCRIBE forecasts had no such opportunity.

There was an anticipated level of performance for each of the three forecast teams (Figure 3). The SCRIBE output was expected to perform the poorly in the earlier part periods of the forecast since it did not use some data sources, such

FORECAST	TODAY	TONIGHT	DAY-2
SCRIBE	3rd	2nd	2nd
Phoenix	1st	2nd	3rd
PSPC	1st	1st	1st

Figure 3. Expected performance of the SCRIBE, Phoenix, and PSPC forecasts during the project.

as radar, and there were inherent programmed restrictions to its performance in the short-term to ensure better performance in the longer term. The Phoenix and PSPC teams were expected to perform well in the short-term since they had access to more data and the traditionally effective short-range forecast techniques.

For the forecasts in subsequent part periods, the SCRIBE forecasts were expected to do better as they were beyond their inherent "spin-up" problems, while the Project Phoenix teams were expected to perform weaker as their short-range and medium range forecast techniques were assumed to be less reliable in this longer time-frame. The PSPC was expected to do the best since those forecasters had access to the full suite of observed data, forecast techniques and numerical guidance.

After the 2-week project, the verification showed the Phoenix and the PSPC meteorologists did handily top the accuracy of the SCRIBE forecasts in the shorter term, with the gap closing rapidly beyond 24 hours (Figure 4). The most intriguing development was that the Phoenix teams managed a significantly better performance than their PSPC counterparts in the shorter terms, suggesting the importance of a greater reliance on data and short-term meteorological techniques.

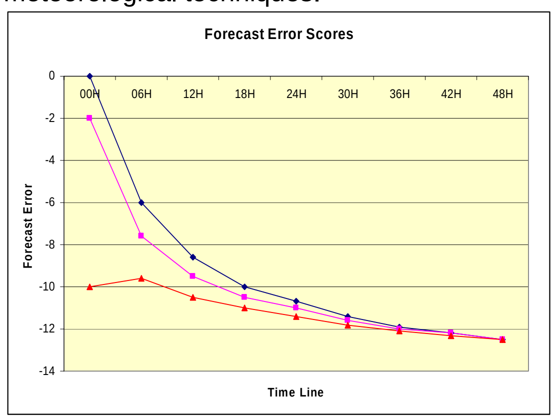

Figure 4. This chart indicates the smoothed error scores for the SCRIBE, PSPC and Phoenix forecasts for nine runs of Project Phoenix. Blue represent the Phoenix team scores, Pink represents the PSPC scores, and red represent s the SCRIBE scores.

In theory, forecasts issued by meteorologists are perfect products at their point of issue. This was the case with the Phoenix forecasts, and nearly so with the official PSPC forecasts, while the automated SCRIBE products were markedly less accurate.

The Phoenix advantage dissipated rapidly, although there was still a marked improvement noted during time intervals of up to 24 hours beyond the issue time of the products. By 48 hours beyond the issue times, there was no gain noted in the Phoenix or PSPC forecasts over the SCRIBE products, although it must be noted that there was virtually no effort made to alter the SCRIBE forecasts beyond this timeframe.

On average, the Phoenix forecast teams reduced the SCRIBE error by 27 percent during the first part-period of the forecasts, while the PSPC achieved a 14 per cent gain over the same computer-generated product. In the second part period of the forecasts, the Phoenix meteorologists produced an 8 percent improvement over SCRIBE, while the PSPC forecasters recorded a 4 percent reduction in error.

Due to the unexpected level of performance by the Phoenix team, a second Project Phoenix was run using three different forecasters. That two-week project and all subsequent Project Phoenixes have yielded similar results.

4. ANALYSIS AND INTERPRETATION OF RESULTS

There were a number of influences that appeared to contribute to the forecast verification results during the Project Phoenix runs.

During each Project Phoenix, participants received daily feedback on performance and the teams routinely discussed issues that appeared to impact their performance. At staff meetings, this discussion expanded to identify a broad range of contributing factors to forecast performance. A PSPC staff was also working with the Canadian Meteorological Centre's numerical weather prediction staff to better understand the factors influencing the SCRIBE performance. Finally, the Project Phoenix coordinator collected this information and performed additional statistical assessments to validate the identified factors. The following summarizes the major performance factors identified:

a. Inertia

Meteorologists were often reluctant to change forecasts significantly, and this had a powerful impact on the forecasts issued by the PSPC. The SCRIBE forecasts had no reluctance, of course, and neither did the Phoenix meteorologist, who generally held the computer-generated forecasts in low regard.

In general, starting with an inferior PSPC forecast resulted in poorer than average error scores throughout succeeding updates, while a superior initial product usually produced better than average error scores in succeeding issues.

b. Minutiae

The Phoenix teams were required to manage its resources in an efficient fashion to produce accurate short-term forecasts, especially if wholesale changes to the SCRIBE forecasts were required. Although the PSPC meteorologists were often reluctant to make big changes to the forecasts prepared by their peers on an earlier shift, they were frequently preoccupied with making minor alterations ("tweaking") to the products, often in the longer-term part periods of the forecasts.

There appeared to be a number of reasons for these changes. Hedging was a common occurrence. "Sunny" or "cloudy breaks" were frequently included, a chance of a shower would be added and a temperature would be tweaked upwards by a degree or two, generally for no apparent meteorological reason. Inevitably, as some of previously recognized, (e.g. Murphy, 1993), such changes resulted in poorer scores more often than not.

Bringing the forecast in line with the finest details indicated by model data was another common reason for the minor changes in the PSPC, especially in subsequent part periods. Minor changes in upper air moisture would prompt a forecast of clearing or clouding over and precipitation timing would be fine-tuned to within a few hours, based on the latest model information. Verification revealed that these detailed changes had a poor chance of success.

A third reason for the trivial PSPC alterations was the use of alternate statistical guidance. Changes of this nature occurred most frequently with temperatures and wind velocities. Once again, verification provided clear evidence that these minor changes were not worth the effort at best, and degraded the original product at least half of the time.

Finally, trivial changes were made when the forecaster would use the model output in lieu of detailed analysis and diagnosis. Early in the Project Phoenix sessions, operational PSPC forecasters, unlike their Project Phoenix counterparts, were observed waiting for the new model data rather the using the time for analysis and diagnosis of observed data. Like the old adage says: “Idle hands are the Devil’s tools”, and the PSPC forecasters often used this “free time” to tweak their forecasts.

Ongoing verification since 2001 has consistently shown that minor changes result in forecast improvements a bit less than 50 percent of the time, adding to forecast error. In fact, when a minor change did bring about an improved forecast, a <u>greater</u> change would have provided further improvement in nearly three-quarters of the cases.

Minor changes to the Phoenix forecasts generally fared no better, but the frequency of minor alterations was markedly lower, especially in subsequent part periods of the forecast. This was due to the lack of temptation offered by a plethora of model data, and that their time was more devoted to analysis and diagnosis and to the accurate predictions of high-impact weather.

c. Competition

The Phoenix teams had a great desire to “beat the model”, as the threat of losing to a machine proved to be a powerful motivator. The competition quickly expanded to include their peers in the PSPC. The competitive spirit within the PSPC also increased, but at a somewhat slower pace over the course of the project.

The competitive nature of some of the Phoenix teams was clearly evident. The meteorologists would report 30 minutes early and review their own performance, cheering their victories and agonizing over defeats. As well, meteorological discussions during the forecast process were generally more animated in the Phoenix office than they were in the PSPC.

The PSPC meteorologists became increasingly competitive during subsequent Phoenix runs. This was likely the result of the consistently better results produced by the earlier Phoenix teams and the fact that the PSPC became increasingly populated with Phoenix veterans.

Many forecasters described the Phoenix experience as “fun”. The Phoenix team environment was usually far more active while the competition aspect generated interest. After initial reservations, many forecasters also enjoyed the thrill of “forecasting without a net” (the models). Generally, forecast teams with an active team dialogue out-performed those teams with individuals who kept more to themselves.

Whatever the motivation, competition consistently produced superior performance, and this trend was observed beyond the Phoenix project. The PSPC was given a target of 25% improvement over the SCRIBE product. With daily feedback, the PSPC performance settled around the 25% mark. When there was no feedback, the scores returned to their pre-Phoenix numbers. If the long term performance number began to slowly slide because of a poor month of performance, forecasters assumed that the still high number meant that they were performing well, rather than they were slipping back into their old habits. When there was a Project Phoenix exercise underway, the PSPC number often dramatically increased above the 25% threshold. In general terms, competition and published verification information improved the routine forecasts in a prediction centre by an average of 12 percent.

d. Experience and abilities

The SCRIBE forecasts are prepared with a constant degree of experience and ability, using the same procedures or rules throughout the period. Although there were staffing changes within the PSPC during the nine Phoenix tests, it would be fair to term the average experience and ability of the PSPC as nearly constant as well.

There were major variations in the experience and the perceived abilities of the Phoenix teams, generally producing the expected variations in performance. Teams composed of more senior and capable forecasters produced the greatest improvements, while the more junior or less capable groups produced products with lower-than-average improvements over the rival forecasts when compared with the Phoenix average.

It is worthy of note that although the poorest Phoenix performance was produced by a team of three recent graduates prior to their first training shifts, they nevertheless managed a significant improvement over the SCRIBE forecasts and their average performance matched that of the PSPC. The same group completed a second Phoenix test a few months later, at the end of their training period, and they produced a lower error score than their PSPC counterparts.

In general terms, although experience was a significant factor, the lack of it was easily mitigated by a willingness to adapt and to adopt new techniques. On the other hand, some experienced meteorologists have struggled within the Phoenix environment.

5. BEYOND PHOENIX – RECOMMENDATIONS AND OUTCOMES

A number of items were identified during the initial Project Phoenix that could lead to improved forecast accuracy within the PSPC, and most of these recommendations were subsequently implemented.

a. Greater use of hand analysis

As noted by others (e.g. Maddox, et al, 1986), hand-analysis/diagnosis generally leads to a greater understanding of the meteorological processes at play, as the meteorologist is forced into a detailed diagnosis and tends to take better ownership of the information at hand. The PSPC has restructured its operation to place added emphasis on analysis, and this appears to have helped improve forecast and warning performance.

(b) Elimination of needless detail

The verification results (Purcell, 2001a, 2001b, 2003, 2004) have consistently underscored the folly of making trivial changes and adding or retaining too much detail in forecasts, especially in the longer term. These problems have declined significantly as a result, resulting in an improved performance and better public perception of the forecast product.

c. Concentrate on the shorter terms

Meteorologists can provide significant gains in accuracy in the shorter timeframes with the use of comparatively minimal effort. Changes in subsequent part periods consume more resources and provide progressively smaller gains. Placing most of the effort on the first part period will therefore achieve the greatest impact in the eyes of the public. The PSPC has managed to boost its verification performance significantly in the first part period of its forecasts, and this gain has carried over into subsequent part periods, where minor gains have also been noted.

d. Use of short-range forecasting techniques

As early as 1970's, Sanders (1979) suggested that over-reliance on NMP in short-term prognosis could compromise the meteorologist's capacity to produce quality forecasts in this time-frame. Phoenix demonstrated that this was indeed that case. However, Phoenix altered this dependence.

In the absence of model and statistical guidance, meteorologists were forced to resort to traditional methods of diagnosis and prognosis. Extrapolation was extensively used in the shortest timeframes and a concerted effort was made to identify nonlinear developments when using the approach in the longer term.

As was noted by John McGinley (1986) and others, there is more to short-range forecasting than simple extrapolation. Additional elements include continuous monitoring and assessment of observed data, a comparison of the subjective analysis with conceptual models both in structure and evolution, and the recognition or anticipation of subtle atmospheric changes critical in the evolution of the phenomenon.

A detailed assessment of the recent weather across the region and how it correlated with the significant weather systems played a major role in the Phoenix process. Contemporary meteorologists, on the other hand, often use this time to evaluate the output of numerical weather prediction. With no such opportunity, the Phoenix meteorologists were able to devote more time to the analysis and diagnosis part of the forecast equation, providing unique forecast alternatives.

The forecast process also benefited from a much greater emphasis on hand analysis of surface and upper air charts. The Phoenix teams also had access to software that objectively analyzed real data, but, as expected (Doswell, 1986b), a hand assessment of the data usually provided a superior result.

In all cases, the short-term techniques proved most effective when changes to the SCRIBE forecasts were limited to the more significant deviations.

e. Real-time and personal verification

The PSPC has routinely conducted real-time verification of its public forecasts since mid-2001 and a significant improvement has been noted in the long-term scores since that time. These improvements are attributed to the added training and awareness provided by the continuation of the Phoenix project, the increased competitiveness brought about by the continued verification, and the added emphasis on analysis and diagnosis using real data.

The PSPC instituted individual verification in 2002. Although individual scores are not officially published, the office makes individual awards on a quarterly basis. An intriguing outcome of this program is that the vast majority of the top performers have less than five years of forecasting experience. This could be due to their willingness and ability to adopt new approaches and a reflection of the difficulty in retraining older staff.

Normally, verification systems should be oblivious to whatever is being tested (Stanksi, 1982). In this case, both Phoenix and PSPC forecasters were aware of the verification system. Only the SCRIBE system was unaffected by the details of the verification scheme. It was important for the forecasters to understand the scheme since it helped them understand the results and it allowed them to identify opportunities to improve their forecasts. Essentially, the verification system was designed to modify their behaviour. Since the verification system measured some level of utility of the forecasts to the public, forecasters could better understand how to improve the “value” of their forecasts to the public. This approach has been controversial and improvements to the scheme continue to be pursued.

Regardless of these concerns, the Project Phoenix verification system has a tremendous impact on the PSPC. There has been a significant change in the culture of the office with respect to verification. Forecasters now routinely discuss performance measurement, the subtleties of the results, ways to glean more information out of this data, and ultimately to share best-practices amoung the group. PSPC staff members are also recognized for their individual forecast performance on a regular basis. Forecasters now demand more sophisticated performance-measurement schemes to help them better understand their performance and to help identify personal training opportunities. More importantly, the forecasters who actively seek personal performance data routinely improve to become amoung the top performers in the office.

f. Case Studies

As a result of the increased awareness brought about by real-time verification, as well as the added emphasis on training and development, the PCPC instituted one-day case studies as a training method. Forecast situations that were particularly challenging are investigated and presented, offering explanations for the forecasts and offering a hindsight look at the situation from a scientific perspective.

Given the massive area of responsibility for forecasters, it is often difficult to identify critical gaps in their training or in the science. With forecasters more focused on the weather that matters most to Canadians, forecasters now more readily identify training and science priorities for themselves and the office. The PSPC has also tried to develop Project Phoenix-based efficiencies in forecast operations to help free up a few additional resources for training and science, as there remains plenty of opportunity for people to add value to the forecasts and warnings through these means.

g. Better use of numerical models

Numerical models remain an important tool for the operational meteorologist. But like any tool, it is important for the forecaster to understand this tool's advantages and limitations. Studies (e.g. Kahn, et al, 2003) have shown that forecasters, who understand their tools well, tend to be the better forecasters.

As Project Phoenix forecasters developed a better situational awareness through their increased analysis and diagnosis, they were able to question the discrepancies of the model output. Sometimes the model information was more correct, sometimes it was not. The forecaster needs to understand why the model handles various situations than others, and thus improving their confidence on making the appropriate changes. This better understanding of model performance will also allow the forecaster to provide more informed feedback and recommendations to model developers.

h. Better tools

One of the common comments from Phoenix participants after they returned to forecast operations was that their operational tools were inappropriate for the job. Forecasters had become increasingly reliant on model information over the years and their tools reflected this. Most workstation software was designed to display and interrogate model gridded binary (GRIB) data, while tools to aid in their analysis and diagnosis of observed data remained neglected, or with the focus on *objective* analysis. PSPC forecasters are now providing insightful recommendations to the developers of Canada's new forecaster workstation (Environment Canada, 2003).

i. New Phoenix approaches

Phoenix has proved to be an effective training approach. All new recruits to the PSPC (now the PAPSC – Prairie and Arctic Storm Prediction Centre) go through a Project Phoenix as indoctrination into the office's forecast philosophy. New forecasters also seem to adopt the approach more readily with most becoming amoung the top performers in the office.

The PSPC now is looking to new Phoenix approaches, including a severe weather Phoenix. Since the original Project Phoenix was designed to test the value of analysis and diagnosis, and by default, the value of numerical weather prediction, similar approaches could test the value of tools such as ensembles, radar, mesoscale models, lightning data, etc., by removing that information from the forecast process. This aspect should be a critical component to a weather event simulator.

The Meteorological Service of Canada's national program to train new recruits is now incorporating aspects of Project Phoenix for their training curriculum.

j. Best practices

The human component of the machine-person mix can break down, compromising the effectiveness of this system. Cognitive task analyses of forecast operations have demonstrated that meteorologists and the operational team must employ "best practices" to be effective (Kahn, et al, 2003).

These best practices include many of the recommendations mentioned in this section. The operational team is a critical aspect of the machine-person mix. The team must optimize the skills of the team, the science of meteorology, their knowledge of the end user, while effectively utilizing all of the tools at their disposal. The team's decision-making must be sound, timely, and effective.

Best practices are becoming a critical component of the Meteorological Service of Canada's training as a result of the lessons gleaned from Project Phoenix.

6. CONCLUSION

Over the past two to three decades, meteorologists have become increasingly reliant upon model guidance, to the point where the plethora of model data available frequently becomes the sole tool used. All too frequently, a meteorologist chooses to resolve a short-term forecasting dilemma by consulting yet another model, seeking "the answer."

The logic of the situation should be clear. It should theoretically be impossible to improve on a model forecast by attempting to reinterpret the same model data, while consulting another individual model for an alternate answer. The project has shown that this approach will yield a poorer result as often as it offers a better one.

The fears of Snellman and others (e.g. Doswell, 1986c, Bosart, 2003) were well-founded. Project Phoenix successfully demonstrated that the skills of PSPC meteorologists had atrophied or had become dormant. However, the project also demonstrated that lost skills could quickly be resurrected and could be used to significantly improve forecast performances.

The project also identified that forecasters have a clear role in the forecast process, by contributing a wealth of knowledge, tools and techniques that can not be duplicated by computers or numerical weather prediction.

Subsequent tests reinforced this result, but also offered tangible proof of the value added by placing a greater emphasis on analysis and diagnosis while using short-term forecasting techniques. The adoption of Phoenix as a training method, combined with continued real-time verification has helped the PSPC maintain and add to its improved forecast accuracy.

Since the system rewards quality high-impact weather forecasts, forecast operations now place a greater emphasis on this aspect. The machine-person mix is more optimized since the technology is focused on the routine and the forecaster is focused on the high-impact weather and where they can appropriately add value. This approach has created new efficiencies within operations. As Doswell (1986a) envisioned, these efficiencies allow forecasters to bring to bear their skills to improve the science, to help enhance technologies, and to improve their knowledge and skills.

The Phoenix approach is now being incorporated throughout the Meteorological Service of Canada. The benefits include better performance, better job satisfaction, better understanding of user needs, more time for meteorologists to be meteorologists, more targeted science and tools development, and better training for future meteorologists. Snellman would be proud.

7. References

Angus Reid Group Inc., 2000. *Public Views on Weather Forecasting Systems*. Prepared for Environment Canada by Angus Reid Group, Inc.

Ball, D., 2007: Project Phoenix: A six year study in back to basics meteorology. *CALMet VII* CD-ROM, Beijing, China, July 2-7, 2007.

Bosart, L. F., 2003: Whither the weather analysis and forecasting process? *Wea. Forecasting*, Amer. Met. Soc., **18**, 520-529.

Boulais, J., 1992: A knowledge based system (SCRIBE/KBS) analyzing and synthesizing large amount of meteorological data. *Preprints, 4th AES/CMOS Workshop on Operational Meteorology*, Whistler, B.C., Sept. 15-18, 1992, 283-286.

_________, G. Babin, D.Vigneux, M. Mondou, R. Parent, and R. Verret, 1992: SCRIBE: an interactive system for composition of meteorological forecasts. *Preprints, 4th AES/CMOS Workshop on Operational Meteorology*, Whistler, B.C., Sept. 15-18, 1992, 273-282.

Decima Research, 2002: National Survey on Meteorological Products and Services – 2002. Final Report. Prepared for Environment Canada.

Doswell, C. A. III, 1986a: The human element in weather forecasting. *Preprints, 1st Workshop on Operational Meteorology,* AES/CMOS*,* Winnipeg, MB, Canada, Feb. 4 - 6, 1985, 1-25.

_________, 1986b: The human element in weather forecasting. *Nat. Weather Digest*., **11**, 6-18.

_________, 1986c: Short-range forecasting. *Mesoscale Meteorology and Forecasting*, edited by Ray, P.S., American Meteorology Society, Boston, Mass., 689-719.

Environment Canada, 2003: National Workstation Project newsletter. Vol. 1, Issue 1, 1-6.

Environics Research Group, 1999. *Evaluation of Precipitation*. Prepared for Environment Canada by Environics Research Group Limited, Toronto, Canada.

Glahn, H. R., 1970: Computer-worded forecasts. *Bull. Amer. Met. Soc.*, AMS, Boston, Mass. 1126-1131.

_________, and D.P. Ruth, 2003: The New Digital Forecast Database of the National Weather Service. *Bull. Amer. Meteor. Soc.*, **84**, 195–201.

Hahn, B. B., E. Rall, D. W. Klinger, 2003: *Cognitive task analysis of the warning forecaster task*. Klein Associates Inc.

Klein, W. H., 1969: The computer's role in weather forecasting. Weatherwise, ***22***. 195-218.

Maddox, R. A., and C. A. Doswell III (1986): The role of diagnosis in operational weather forecasting. Preprints, *11th Conf. on Weather Forecasting and Analysis*, Kansas City, MO, American Meteorological Society, 177-182.

Mass, C.F., 2003: IFPS and the Future of the National Weather Service. *Wea. Forecasting*, **18**, 75–79.

McGinley, J., 1986: Nowcasting mesoscale phenomena. *Mesoscale Meteorology and Forecasting*, edited by Ray, P.S., Amer. Meteor. Soc., Boston, Mass., 657-688.

Murphy, A.H., 1993: What makes a good forecast? An essay on the nature of goodness in weather forecasting. *Wea. Forecasting*, **8**, 281-293.

Purcell, W. M. 2001a: *Project Phoenix – Preparing Meteorologists for an Intensive Man-Machine Mix*. Environment Canada, Meteorological Service of Canada, 1-15.

___________ 2001b: *Project Phoenix II – Preparing Meteorologists for an Intensive Man-Machine Mix.* Environment Canada, Meteorological Service of Canada, 1-10.

Rezek, A., 2002: Changing the Operational Paradigm with Interactive Forecast Preparation System – IFPS. Preprints, *Interactive Symposium on the Advanced Weather Interactive Processing System AWIPS),* Orlando, FL, Amer. Meteor, Soc., 23–27.

Ruth D. P., 2002: Interactive forecast preparation—the future has come. Preprints, *Interactive Symp. on the Advanced Weather Interactive Processing System (AWIPS),* Orlando, FL, Amer. Meteor. Soc., 20–22.

Sanders, F., 1979: Trends in Skill of Daily Forecasts of Temperature and Precipitation, 1966–78. *Bull. Amer. Meteor. Soc.*, **60**, 763–769.

Snellman, L.W., 1977: Operational forecasting using automated guidance. *Bull. Amer. Meteor. Soc*., **58**, 1036-1044.

Stanski, H., 1982: Module VIII – Verification of Forecasts. *AES Training Branch Correspondence Course.* Environment Canada.

__________, L.J. Wilson, and W.R. Burrows, 1989: *Survey of common verification methods in meteorology.* World Weather Watch Tech. Rept. No.8, WMO/TD No.358, WMO, Geneva, 114 pp.

Verret, R., G. Babin, D. Vigneux, R. Parent, and J. Marcoux, 1993: SCRIBE: an interactive system for composition of meteorological forecasts. *Preprints, 13th Conf. on Weather Analysis and Forecasting*, Vienna, Virginia, August 2-6, 1993, Amer. Met. Soc. 213-216.

________, G. Babin, D. Vigneux, J. Marcoux, J. Boulais, R. Parent, S. Payer and F. Petrucci, 1995: SCRIBE: an Interactive system for composition of meteorological forecasts. *Preprints, 5th AES/CMOS Workshop on Operational Meteorology,* Edmonton, AB, Canada, Feb. 28 - March 3, 1995, 33-39.