Verification of Spot Fire Weather Forecasts
Project ID: 12-1-05-3

Principal Investigator: John D. Horel, University of Utah, Department of Atmospheric Sciences
Co-I: Timothy Brown, Desert Research Institute

Abstract

Software was developed to evaluate National Weather Service (NWS) spot forecasts. Fire management officials request spot forecasts from the NWS to provide detailed guidance as to atmospheric conditions in the vicinity of planned prescribed burns as well as wildfires that do not have incident meteorologists on site. A multi-year set of spot forecasts of maximum temperature, minimum relative humidity, and maximum wind speed were compared to nearby surface observations as well as gridded values from the NWS National Digital Forecast Database (NDFD). Based on spot forecasts nationwide, their skill is higher than that available from the NDFD, with the greatest improvement for maximum temperature (8-10% improvement) and less so for maximum wind speed (2% improvement). Verification using nearby soundings of mixing height, transport wind and Haines Index forecasts indicated that mixing height forecasts exhibited larger errors and tended to be biased towards overforecasting compared to forecasts of transport wind speeds and Haines Index values.

An overarching recommendation of this study is to leave the decisions as to what to verify and how to verify the forecasts in the hands of the forecasters and end users by developing flexible methods to explore the multidimensional nature of the forecasts. Based on qualitative, in-depth interviews with fire practitioners and NWS forecasters, improving accuracy and utilization of spot forecasts requires improving communication between NWS Forecast Offices and fire personnel in the field. In addition, study participants recommended considering to provide forecast uncertainty and forecaster confidence in spot weather forecasts to open the opportunity for fire practitioners to use this information in risk analysis. A number of specific recommendations that would increase the usefulness of spot forecasts include: (1) increase the consistency of the information provided by the spot forecasts, (e.g., isolate quantitative numerical values separately from qualitative alphabetical descriptors); 2) assemble a sizable sample of focused prescribed fire and wildfire case studies to evaluate and verify the forecasts in greater depth; 3) improve the spot request process by allowing the requestor to provide information pertinent for verification (e.g., requested forecast wind level a numerical parameter adjustable by the end user within the request form as well as potential nearby observation locations to compare to the forecast).

A noteworthy aspect of this research has been to facilitate transfer from research to operations of the techniques and web-based tools used to undertake the spot forecast verification (see http://meso1.chpc.utah.edu/jfsp/). Approaches used in this study to verify spot forecasts are being migrated to the operational environment of the NWS Performance and Evaluation Branch.

Background and Purpose

A 2008 National Oceanic and Atmospheric Administration (NOAA) report entitled, "Fire Weather Research: A Burning Agenda for NOAA," outlined the need for more robust forecast verification for wildland fire incidents. National Weather Service (NWS) forecasters at Weather Forecast Offices (WFOs) have issued 103 370 forecasts, often at very short notice, requested by fire and emergency management professionals for specific locations, or "spots", during the April 2009–November 2013 period. Spot forecasts are requested for prescribed burns, wildfires, search and rescue operations, and hazardous material incidents. NWS forecasters rarely receive detailed feedback from fire and emergency management professionals on the usefulness of their spot forecasts and no quantitative evaluation of spot forecasts has been undertaken nationwide.

Prescribed fires on federal or state land have operating plans that contain thresholds for atmospheric variables such as wind speed and relative humidity beyond which they should not commence burning. Spot forecasts play a central role in determining whether a burn is initiated on a given day. Of the 16 600+ prescribed burns undertaken in 2012, only 14 escaped. However, public reaction to this small number of escapes is overwhelmingly negative. Outcry from the Lower North Fork Fire, which broke out in smoldering litter four days after the prescribed burn work, destroyed 23 homes, caused three fatalities and led to modifications of the Colorado state constitution to allow victims of prescribed burn escapes to sue the state. The nation is increasingly at risk for loss of life and damage to property as a result of wildfires. During 2003, fires near San Diego, California destroyed over 3500 homes and killed 22 people. Three fires (High Park, Waldo Canyon, and Black Forest) in the Front Range of Colorado in 2012 and 2013 destroyed a total of 1117 homes.

Forecast guidance helps to determine the magnitude and placement of responding firefighters. Guidance is issued by WFO forecasters initially and later by Incident Meteorologists as wildfires grow in extent. In some circumstances, there is little that can be done to contain explosively developing conflagrations, but even when the ability to control a fire is diminished, accuracy in forecasting the timing and intensity of fire growth is essential. The deaths of 19 firefighters in Yarnell, Arizona, caused in part by a sudden wind shift outflowing from a thunderstorm, underscore the need for addressing the wide range of possible fire weather conditions in forecasts.

This research was focused on verifying spot forecasts using administrative (assess overall forecast performance for strategic planning) and scientific (improve understanding of the nature and causes of forecast errors to improve future forecasts) approaches. We used both measures-oriented and distributions-oriented verification techniques. The former is centered on statistics such as bias, root-mean squared error, or skill scores developed to contrast forecasts with verifying data. The distributions-oriented method presents more detailed information about the relationships between the forecasts and the verifying observations using joint, marginal, and conditional distributions to inspect categorical errors as a function of other factors.

The objective verification was done in two parts with separate verification of key variables at the surface (temperature, moisture, and wind) and aloft (mixing height, transport winds, and Haines Index). In addition, a separate study evaluating the utility of spot forecasts was completed based on feedback from forecasters and fire personnel who rely on those forecasts. The principal results obtained from all three studies are summarized below.

Study Description

This study was designed to overcome a number of limitations imposed by the way the thousands of spot forecasts are created and disseminated, e.g.,: the mix of textual and numerical values contained in spot forecasts makes it difficult to extract pertinent information for verification and the numerical values contained within the spot forecasts are not separated and sent to a centralized online database. Natural language methodologies were developed as part of this project to parse the forecast values from the freeform text of the spot forecasts. Figure 1 shows the number of spot forecasts available to be analyzed for prescribed (top) and wild (bottom) fires from 2009 to the present day (August 25, 2015).

This study developed procedures to access, archive, and display the spot forecasts, the observations and analysis values used to verify the forecasts, and the resulting verification statistics. In order to be able to rapidly query such a large dataset that is continually updating, a comma-separated text file containing every valid forecast with the corresponding nearby observations, NDFD forecasts, and RTMA values has been created and continues to do so. To alleviate the complexity of the multivariate nature of the spot forecasts, the open source Crossfilter code developed by Square, Inc., is used that allows for near-instantaneous slicing on each axis of a multidimensional data set. That allows users to create histograms conditioned on ranges of values in multiple dimensions, i.e., within selected elevation ranges, times of year, values of variables. These histograms then can be adjusted dynamically by the user based on selections in other histograms. The Crossfilter object is instantiated by simply pulling in the necessary information in comma-separated format. Filters are generated on one or more of the variables so that the user can make selections based on ranges of values, but also visualize the impact of other selections on these variables. For further details, see Lammers and Horel (2014) or the web page developed for this project: http://meso1.chpc.utah.edu/jfsp/.

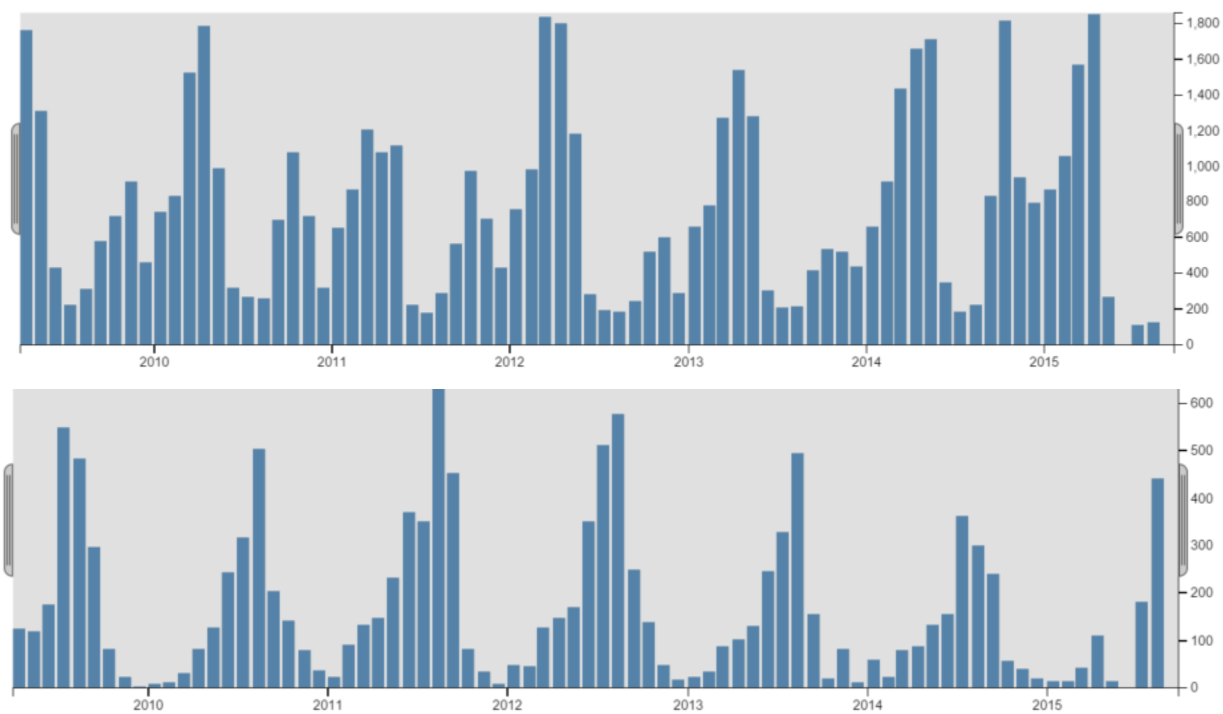

Figure 1. Number per month of prescribed fires (top) and wildfires (bottom) that can be analyzed from 2009 through the present day using the software available from http://meso1.chpc.utah.edu/jfsp/.

In addition, spot forecast requests placed within 50 km of an atmospheric sounding location were organized and saved by year (Figure 2). Text and numerical values associated with mixing height, transport winds, and Haines Index were extracted from spot forecasts valid the same day they were issued within that distance from a sounding location. The calculated variables from atmospheric soundings were directly compared to the spot forecast numerical values. If there was one forecast value, mean errors (MEs) and mean absolute errors (MAEs) were calculated for each of the applicable calculated variables. If the spot forecast issued a forecast range and an atmospheric sounding calculated value occurred between the lower and upper bound of that forecast range, the ME and MAE was zero. If the calculated variable did not occur within the forecast range, then it was compared to the closest value, either the lower or upper bound, with ME and MAE calculated. Consideration was given to implementing an acceptable error range for single value forecasts, such as plus or minus five percent of the forecast value, which could have alleviated some of the bias towards range forecasts.

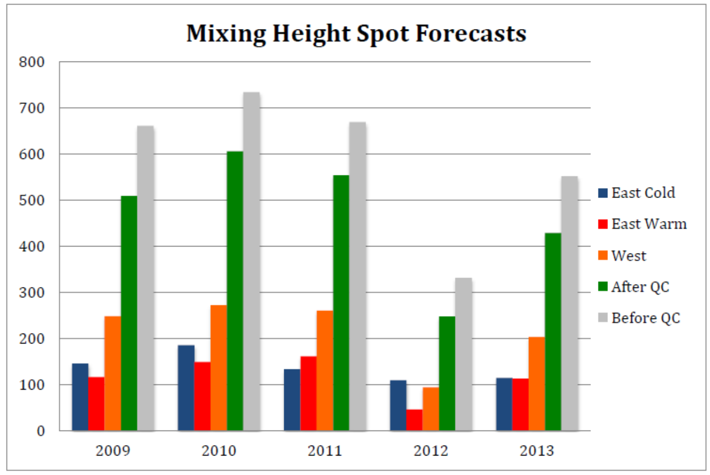

Figure 2. Number of mixing height spot forecasts evaluated by year and category used in the study.

Finally, this research project used qualitative, in---depth, semi---structured interviews with focus groups and individuals. There were nine focus group interviews, with approximately 90 students attending an S---490 course (Advanced Fire Behavior Calculations) in three geographical areas of the United States in 2014. The S---490 course is attended by firefighters who are selected based on their service record and desire to be qualified as a fire behavior analyst, long---term fire analyst, prescribed fire burn boss or deemed to benefit from completion of the course. An initial analysis of data from the focus group interviews was presented to a group of NWS forecasters and fire weather/behavior experts. Based on their suggestions, we conducted a smaller and more focused second round of phone interviews in 2015 with seven NWS forecasters. The geographic distribution of these interviews matched that from the first round of interviews. We prepared a semi---structured interview guide for both the fire practitioners and forecasters. Interviews were generally 45---60 minutes in length. The forecasters were selected from within their geographic region by a snowball sampling method. All of the interviews were recorded, transcribed, and coded. Emergent themes were used to synthesize data in a second round of coding in thematic groups. Our final analysis focused on developing and expanding relationships between the thematic groups.

Key Findings

Verification of surface weather elements of spot forecasts

Software was developed to evaluate National Weather Service spot forecasts. Fire management officials request spot forecasts from National Weather Service Weather Forecast Offices to provide detailed guidance as to atmospheric conditions in the vicinity of planned prescribed burns as well as wildfires that do not have incident meteorologists on site. This open source software with online display capabilities is used to examine an extensive set of spot forecasts of maximum temperature, minimum relative humidity, and maximum wind speed from April 2009 through November 2013 nationwide. The forecast values are compared to the closest available surface observations at stations installed primarily for fire weather and aviation applications. The accuracy of the spot forecasts is compared to that available from the National Digital Forecast Database (NDFD).

Spot forecasts for a selected prescribed burn are used to illustrate issues associated with the verification procedures. Cumulative statistics for National Weather Service County Warning Areas and for the nation are presented. Basic error and accuracy metrics for all available spot forecasts and the entire nation indicate that the skill of the spot forecasts is higher than that available from the NDFD, with the greatest improvement for maximum temperature and the least improvement for maximum wind speed.

Verification of upper-air weather elements of spot forecasts

Fire management officials request spot forecasts from NWS Weather Forecast Offices (WFOs) to provide detailed guidance of atmospheric conditions in the vicinity of prescribed and wildland fires. Verifying spot forecasts represents an integral component of the forecast process and helps assess and improve the accuracy of forecasts. The verification analysis here utilized NWS spot forecasts of mixing height, transport winds and Haines Index from 2009-2013 issued for a location within 50 km of an upper sounding location and valid for the day of the fire event. Mixing height was calculated from the 0000 UTC sounding via the Stull, Holzworth, and Richardson methods. Transport wind speeds were determined by averaging the wind speed through the boundary layer as determined by the three mixing height methods from the 0000 UTC sounding. Haines Index was calculated at low, mid, and high elevation based on the elevation of the sounding and spot forecast locations. Forecast statistics were calculated for each boundary layer element by region including mean error and mean absolute error. Mixing height forecasts exhibited large mean absolute errors and biased towards over forecasting. Forecasts of transport wind speeds and Haines Index exhibited relatively smaller mean error and median absolute error than mixing height forecasts.

Improving utilization, communication, and perceptions of the accuracy of spot forecasts

Spot Weather Forecasts are issued by Weather Service Offices throughout the United States, primarily for use by wildfire and prescribed fire practitioners for monitoring local---scale weather conditions. This paper focuses on the use of Spot Weather Forecasts by prescribed fire practitioners to assess fire weather conditions. Based on qualitative, in---depth interviews with fire practitioners and National Weather Service forecasters, this study examined factors that influence both quantitative accuracy as well as perceptions of accuracy, and how accuracy and utilization of Spot Weather Forecasts can be improved. Results indicate that several well understood climatological, topographical, and data---driven factors influence forecast accuracy. A key opportunity for improving accuracy and utilization of these forecasts, however, may be in enhancing the process and mechanisms for communication between a Weather Forecast Office and fire personnel in the field. In addition, study participants identified a need to consider including forecast uncertainty and forecaster confidence in Spot Weather Forecasts, the challenges for forecasters to do so, and the possibility of fire practitioners using this information in risk analysis.

Management implications

Forecast verification is a continual, ongoing process. Tools must be in place that make it possible for the forecasters and users of the forecasts to quickly examine cases and aggregate statistics of interest to them using their experience and local knowledge, rather than depending on bulk statistical metrics accumulated on national scales. In order to develop useful verification tools for spot forecasts operationally requires minimizing some of the underlying limitations identified during this study. **The principal recommendation of this study is to leave the decisions as to what to verify and how to verify the forecasts in the hands of the forecasters and end users by developing flexible methods to explore the multidimensional nature of the forecasts.** Foremost is simply the need to be able to examine in a centralized framework: the requests; the forecasts; geolocation information; and nearby observations and other information relevant to analyzing the forecasts. Then, the user should be able to explore and control interactively key parameters (e.g., distance to the verifying observations, forecast lead times, magnitudes of the parameters, or magnitudes of the errors). Currently, much of the verification performed on the federal level boils down to aggregate statistics that fail to capture the nuance necessary for evaluating spot forecasts, something that the online tools enable. In order to make the tools described in this study more appropriate for operational use, several limitations need to be overcome:

- Isolate quantitative numerical values separately from qualitative alphabetical descriptors.
- Make forecast wind level a numerical parameter adjustable within the request form, so that even when it is not "20-Foot," the level is known for evaluation.
- Store the name of or abbreviation referencing the specific station for verification as part of the request form. This should include stations from networks not used in this study.

Additionally, the results of this spot forecast verification highlights some other recommendations for any future operational attempt to evaluate spot forecasts, such as planned by the NWS Performance Branch:

- The consistency of the information provided by the spot forecasts needs to be improved.
- A framework for verification of spot forecasts needs to be developed and implemented. Without separating numerical content or increasing the standardization of spot forecasts, any verification method implemented will encounter the inconsistencies and ambiguity in spot forecasts, which will mitigate the verification's potential positive impact.
- Assemble a sizable sample of focused prescribed fire and wildfire case studies to evaluate and verify forecasts. Examining the forecasts made during these prescribed burns and wildfires provide insight into possible sources of consistent errors that may lead to improving forecasts
- Establish accuracy thresholds or requirements of spot forecasts. This would engage the user community and provide an opportunity for NWS forecasters and users to communicate issues related to spot forecast performance.

Results from our research also suggest that the issues surrounding communication between forecasters and fire practitioners are equally important in both the perceptions of accuracy and quantitative accuracy. These issues can be summed up largely as follows: (1) communication challenges before, during, and after an SWF is issued and (2) communication of uncertainty and confidence in the forecast. When an SWF is issued, the communication that took place between the WFO, forecasters, and fire practitioner community is often an essential precursor to the forecast's perception as credible and salient by the user. During a prescribed fire, the primary challenges are updating an SWF if changes in weather conditions critical to fire behavior occur and communication efficacy between forecasters and prescribed fire personnel. Our results suggest that the process of monitoring weather conditions and communicating changes, as two---way communication between the WFO and field personnel, is limited without well defined protocols. In addition, there are workload issues within the WFOs' and forecasters' abilities to effectively monitor SWFs on high volume days. After a prescribed fire, or on multi---day burns, providing feedback on SWF accuracy faces several barriers. Fire practitioners – already burdened with post---burn obligations such as monitoring, releasing resources, equipment cleaning, and paperwork – are often not incentivized to provide feedback to forecasters. This may be especially true if the forecast were accurate, but as several forecasters noted, they need feedback for both inaccurate and accurate SWFs to effectively improve accuracy. Fire practitioners also may be reluctant to "criticize" forecasters for what seem like small errors but, cumulatively, may have a perceptible influence on accuracy.

Transfer of Research to Operations

A significant aspect of this research has been to facilitate transfer from research to operations of the techniques and web-based tools used to undertake the spot forecast verification (see http://meso1.chpc.utah.edu/jfsp/ and Figure 3). Approaches used in this study to verify spot forecasts are being migrated to the operational environment of the NWS Performance and Evaluation Branch.

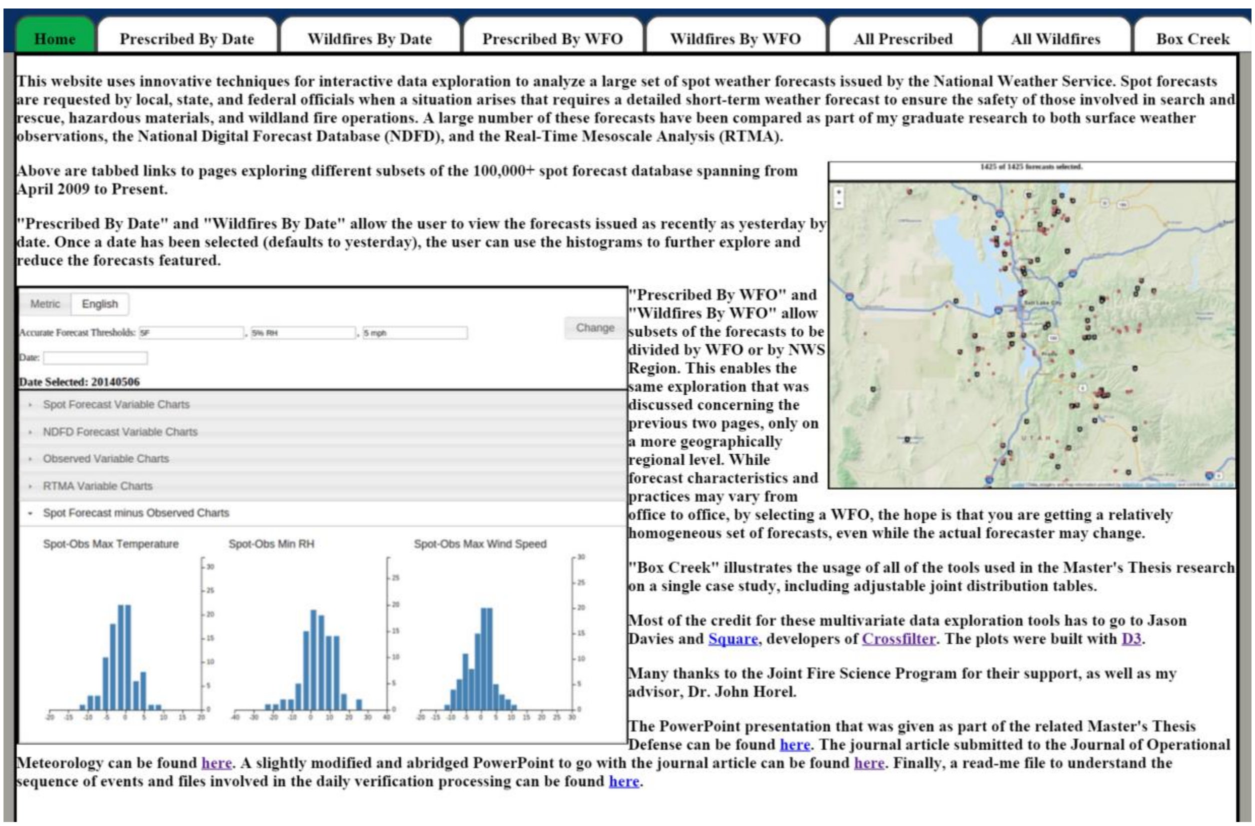

Figure 3. Landing page of http://meso1.chpc.utah.edu/jfsp/

Future work needed

There is a clear need to establish accuracy thresholds or requirements for spot forecasts. This would engage the user community and provide an opportunity for NWS forecasters and users to communicate concerning spot forecast performance. Current JFSP funded work is examining aspects of accuracy concerning weather data in the context of management decision-making, which could help address this issue.

The transition of research to operations needs to continue with implementation of the verification approaches developed here by the NWS Performance Branch. However, that system requires users to have an account, which is automatic for NWS personnel but not for fire professionals. A more open verification system accessible by all without accounts should be the end goal for general information on forecast accuracy and uncertainty.