

Verification of National Weather Service Spot Forecasts Using Atmospheric Sounding Observations

NICHOLAS J. NAUSLAR and TIMOTHY J. BROWN
Desert Research Institute, Reno, Nevada

JOHN D. HOREL
University of Utah, Salt Lake City, Utah

(Manuscript received 8 July 2015; review completed 19 January 2016)

ABSTRACT

Fire management officials request spot forecasts from National Weather Service (NWS) Weather Forecast Offices to provide detailed guidance regarding 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 utilizes NWS spot forecasts of mixing height, transport winds, and the Haines index (HI) from 2009–2013 issued for a location within 50 km of an upper-air sounding site 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. The HI was calculated at low, mid, and high elevation based on the elevation of the sounding and spot-forecast locations. Forecast statistics, including mean error and mean absolute error, were calculated for each lower-atmospheric variable by region. Mixing height forecasts exhibited large mean absolute errors and were biased toward overforecasting. Forecasts of transport wind speeds and HI outperformed mixing height forecasts with smaller errors relative to their respective means. Based on these results and the methodology, recommendations are provided to improve spot forecasts and the verification process.

1. Introduction

A 2007 report entitled “National Wildland Fire Weather: A Summary of User Needs and Issues” from the Office of the Federal Coordinator for Meteorological Services and Supporting Research (OFCM) emphasized a number of improvements that are needed, including that the fire community should “establish accuracy requirements for fire weather products and services to enable the provider community to focus improvement efforts where most beneficial” (OFCM 2007). OFCM (2011) updated the responses to these findings but indicated that the original findings still had not been addressed adequately. A 2008 National Oceanic and Atmospheric Administration (NOAA) report entitled, “Fire Weather Research: A Burning Agenda for NOAA,” echoed similar sentiments by identifying the need for improving and conducting a more thorough forecast verification for wildland fire

incidents (NOAA Science Advisory Board 2008; Lammers and Horel 2014).

National Weather Service (NWS) forecasters at Weather Forecast Offices (WFOs) issue spot forecasts in response to requests from fire and emergency managers. These forecasts are typically for prescribed fire, wildfire, hazardous material, and search-and-rescue incidents and provide detailed guidance for atmospheric conditions in the vicinity of these incidents. The NWS issues approximately 20 000 spot forecasts each year for prescribed fires and wildfires, which comprise the vast majority of the spot forecasts. The NWS issues prescribed fire spot forecasts nearly twice as often as wildfire spot forecasts (Lammers and Horel 2014).

Brier and Allen (1951), Joliffe and Stephenson (2003), and Wilks (2011) identified and demonstrated appropriate verification techniques to assess forecast

performance and to understand sources of error to improve future forecasts. Brown and Murphy (1987) detailed a fire weather forecast evaluation process that identifies biases based on forecasters' perceived consequences of underforecasting key fire weather variables and the difficulties of quantifying uncertainty in their forecasts.

Lammers and Horel (2014) examined spot forecasts from April 2009 through November 2013 and evaluated spot forecasts of surface temperature, moisture, and wind by using surface observations from the closest surface station (e.g., remote automated weather stations) and the National Digital Forecast Database (NDFD). Spot forecasts demonstrated higher skill than NDFD output and improved upon NDFD output for all examined forecast elements, especially for maximum temperature—whereas the smallest improvement was associated with maximum wind speed. Our paper expands on this previous work by evaluating spot forecasts of mixing height (MH), transport winds (TWs), and Haines index (HI).

There were two primary objectives for the analysis.

- 1) Attempt to objectively verify MH, TWs, and HI in spot forecasts.
- 2) Demonstrate and review the spot verification process and offer relevant recommendations to improve spot forecasts.

Section 2 lists the data utilized in this study, and section 3 describes the methods implemented to perform the spot-forecast verification. Section 4 presents the results, and sections 5 and 6 discuss the results and draw conclusions from the current research, including noting important caveats and offering recommendations.

2. Data

For the period 2009–2013, 89 052 NWS spot forecasts were initially gathered for analysis (Table 1). Spot-forecast requests contain the date, NWS WFO, the incident's name, latitude, longitude, elevation, the forecast parameters needed, and the option to select “today,” “tonight,” and “tomorrow” for when those forecasts should be valid (Fig. 1). The spot forecast contains a short narrative of the weather forecast and then lists values for each forecast element and when they are valid (Fig. 2). For example, a spot forecast

might contain three different forecasts: “today,” “tonight,” and “tomorrow.”

Many definitions and names describe the planetary boundary layer, including boundary layer (BL) and mixed layer (ML). Stull (2000) described the BL as the shallow layer near the surface where the diurnal variation of sensible and latent heat fluxes exists between the surface and atmosphere. Wallace and Hobbs (2006) defined the BL as the layer most affected by the earth's surface, which is separated from the rest of the troposphere because of the effects of turbulence and static stability. The BL undergoes diurnal variation, but the variation changes depending on a number of factors, including season, terrain, synoptic conditions, and land-surface type. Typically, the shallowest BL occurs just before sunrise, and as radiation flux increases, the BL builds throughout the day, peaking in height during the afternoon (Fig. 3; adapted from Stull 2000).

A general definition of MH is the top of the BL or ML. However, there is no universally accepted definition or criteria for its determination for two reasons: 1) various processes—such as turbulence, radiation, baroclinicity, advection, divergence, convergence, and vertical motions—contribute to the structure of the BL; and 2) most definitions or criteria are constructed on available data measured through various instruments and techniques (Beyrich 1997). Within the BL or ML, the mean wind speed and direction are defined as the TWs (Miller 1967; AirFire 2015; NWS 2015). AirFire (2015) noted that some state and local agencies vary their definition of TWs, including using a weighted mean for wind speed and direction through the BL. The HI qualitatively assesses dry, unstable air present in the lower- to mid-troposphere, which can increase wildfire behavior (Haines 1988; Werth et al. 2011). Daily fire weather forecasts utilize the HI for determining fire potential, especially for plume-dominated wildfires (Haines 1988; Werth and Ochoa 1993; Potter et al. 2008; Werth et al. 2011).

Atmospheric sounding data (obtained online from weather.uwyo.edu/upperair/sounding.html) from 2009 to 2013 and valid at 0000 UTC were selected on fire days and nearest to fire locations. Variables retrieved from the sounding included unit identifier, latitude, longitude, elevation, potential temperature (θ), virtual potential temperature (θ_v), wind speed, pressure, and height. These variables were necessary to calculate the three BL elements examined for comparison to spot forecasts.

Table 1. Total number of spot forecasts issued each year and number of spot forecasts issued within 50 km of an atmospheric sounding.

	2009	2010	2011	2012 (through May)	2013 (through August)	Total
All forecasts	22 077	20 846	21 678	9918	14 533	89 052
Forecasts within 50 km of a sounding	1365	1347	1364	628	1063	5767

RENO SPOT FORECAST REQUEST

Required Elements in RED (*)

PROJECT NAME

(*)Project Name:

Wildfire HAZMAT
 Prescribed Fire SAR
 Ignition Time: Pacific Local Time
 Date:

REQUESTING AGENCY

(*)Requesting Agency:

(*)Requesting Official:

(*)Phone Number: Ext.

FAX Number:

Contact Person:

REASON FOR SPOT FORECAST REQUEST

(*)Must choose either Wildfire or one of the Non-Wildfire reasons

Wildfire

Non-Wildfire

Under the Interagency Agreement for Meteorological Services (USFS, BLM, NPS, USFWS, BIA).
 State, tribal or local fire agency working in coordination with a federal participant in the Interagency Agreement for Meteorological Services.
 Essential to public safety, e.g. due to the proximity of population centers or critical infrastructure.

For NWS Spot forecast policy, see section 4.0 in NWS Instruction 10-401 at <http://www.nws.noaa.gov/directives/010/010.htm>

LOCATION

(*)Lat:

(*)Lon:

7.5' Quad:

Legal (T/R):

NV
 CA

ELEVATION

(*)Elevation: Top Bottom

Drainage:

(*)Aspect:

Size: (Acres)

OBSERVATIONS

Place	Elev	Time	Wind	Temp	Wetbulb	RH	Dewpt.	Sky/Weather
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

FUEL

Type:

Sheltering
 Full
 Partial
 Unsheltered

PRIMARY FORECAST ELEMENTS

TDA TNT TMR (Today, Tonight, Tomorrow)

Clouds / Weather
 Temperature
 Relative Humidity
 20 Foot Wind
 Smoke Dispersion

REMARKS

Figure 1. An example of the spot-forecast request form used by the NWS WFO Reno, NV. *Click image for an external version; this applies to all figures hereafter.*

3. Methods

a. Spot forecasts

Spot-forecast requests placed within 50 km of an atmospheric sounding location were organized and saved by year. Lammers and Horel (2014) chose 50 km as an appropriate distance when comparing spot forecasts and surface station observations. This verification only used the “today” forecasts from spot forecasts containing MH, TWs, and HI. In summary, only spot forecasts issued and valid on the same day and within 50 km of an atmospheric sounding were considered.

Each spot-forecast request corresponds to a spot forecast. The spot-forecast request contains the date, incident name, NWS WFO, and other metadata that are used to find the corresponding spot forecast, which contains all of the forecast data that are verified. Multiple spot-forecast requests and corresponding spot forecasts could be associated with the same incident. Spot forecasts are requested daily for many wildfires until an incident meteorologist arrives to handle forecast responsibilities. In addition, burn bosses for prescribed fires often request consecutive daily spot forecasts to ascertain if the weather conditions permit lighting the prescribed fire. If the spot forecasts for the

```

FNUS75 KGJT 210312
FWSGJT

SPOT FORECAST FOR FRANKS BENCH...GRAND JUNCTION DISPATCH
NATIONAL WEATHER SERVICE GRAND JUNCTION CO
912 PM MDT TUE AUG 20 2013

FORECAST IS BASED ON REQUEST TIME OF 1850 MDT ON AUGUST 20.
IF CONDITIONS BECOME UNREPRESENTATIVE...CONTACT THE NATIONAL WEATHER
SERVICE.

.DISCUSSION...A PLUME OF SUB TROPICAL MOISTURE SITTING OVER THE
REGION WILL RESULT IN ANOTHER ROUND OF SHOWERS AND THUNDERSTORMS
WEDNESDAY AFTERNOON AND EVENING. STORMS WILL DIMINISH LATE IN THE
EVENING...HOWEVER A FEW SHOWERS WILL CONTINUE TO BE POSSIBLE
OVERNIGHT. STORMS WILL BE SLOW MOVING WITH HEAVY RAIN POSSIBLE IN
THE AFTERNOON AND EVENING.

.WEDNESDAY (UNTIL SUNSET)...

SKY/WEATHER.....PARTLY CLOUDY (50-60%)...BECOMING MOSTLY
CLOUDY (70-80%) IN THE AFTERNOON. AN ISOLATED
SHOWER OR THUNDERSTORM POSSIBLE IN THE MORNING...
THEN SCATTERED SHOWERS AND THUNDERSTORMS
DEVELOPING IN THE AFTERNOON.

MAX TEMPERATURE.....80-85 F.
MIN HUMIDITY.....28-33%.
20-FOOT WINDS.....DOWNSLOPE/DOWNVALLEY 3-6 MPH UNTIL 1000...THEN
LIGHT AND VARIABLE. BECOMING UPSLOPE/UPVALLEY
4-8 MPH AFTER 1200. GUSTY AND ERRATIC WINDS
EXPECTED NEAR THUNDERSTORMS. BY LATE AFTERNOON
THUNDERSTORM OUTFLOW WINDS MAY DOMINATE
SURFACE AND 20-FOOT WINDS WITH GUSTS TO
35 MPH.

SMOKE DISPERSAL.....POOR UNTIL 1100...THEN BECOMING GOOD IN THE
AFTERNOON.

LAL.....2 UNTIL 1200...THEN 4.
MIXING HEIGHT.....INCREASING TO 11000 FT AGL IN THE AFTERNOON.
TRANSPORT WINDS.....WEST 5 MPH. VARIABLE THUNDERSTORM OUTFLOWS MAY
DOMINATE TRANSPORT WINDS AT TIMES.
    
```

Figure 2. Example of a spot forecast issued by NWS WFO Grand Junction, CO.

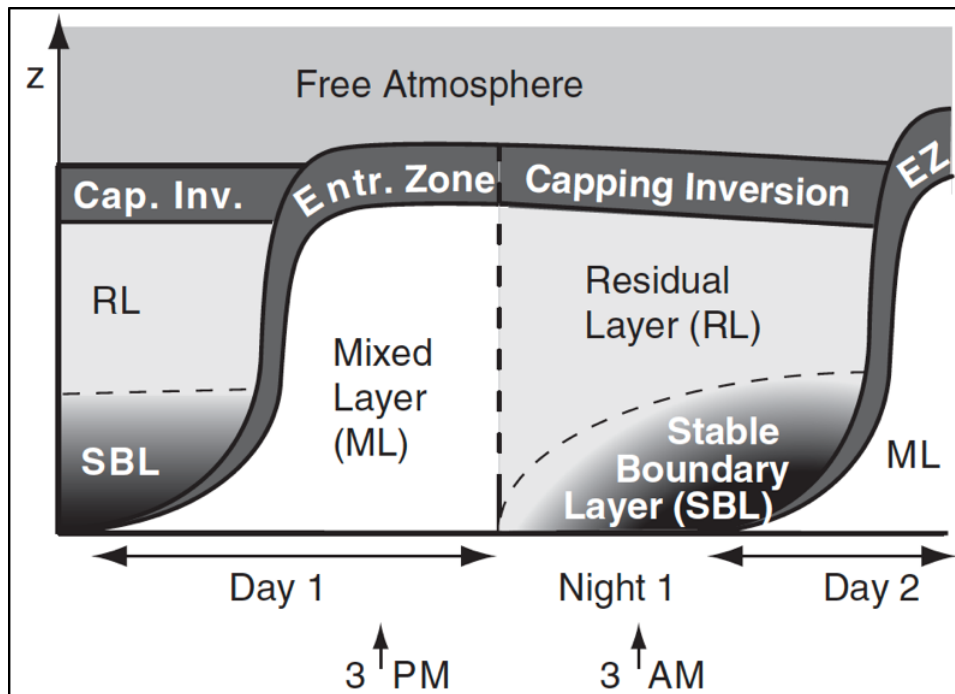


Figure 3. Diurnal boundary layer transition vertical profile from Stull (2000).

same incident transpired on different days, then the spot forecasts were included in the analysis. If NWS WFOs issued more than one spot forecast on a particular day for an incident, only the most recent spot forecast was used in the analysis.

b. Atmospheric soundings

NWS WFOs launch the 0000 UTC atmospheric sounding around 2300–2315 UTC. Table 2 shows the local times for launching the rawinsonde of the 0000 UTC atmospheric sounding. The -102.5° longitude line, which roughly parallels the central and mountain time zone border, is used to separate spot forecasts and their corresponding atmospheric soundings into west and east categories. The east category was then split into an east warm season (EWS) category spanning April–October and an east cold season (ECS) category containing the remaining months. Because 0000 UTC occurs during the afternoon or early evening throughout the western United States, the nocturnal BL should not affect the MH calculation. The time period 2300–0000 UTC occurs in the early evening for most of the central and eastern United States during the warm season (Table 2). However, during the cold season, most of the central and eastern United States are near or after sunset when 2300 or 0000 UTC passes (Table 2). Splitting into west and east categories and subdividing the east category into ECS and EWS categories mitigates some of the potential spot-forecast verification problems with the nocturnal BL and isolates potential inapplicable comparisons.

Table 2. Atmospheric sounding launch times for each time zone during standard and daylight times (for the 2300 UTC standard launch).

Time Zone	Launch (standard time)	Launch (daylight time)
Eastern	1900	1800
Central	1800	1700
Mountain	1700	1600
Pacific	1600	1500

c. Parsing spot forecasts and atmospheric soundings

Lammers and Horel (2014) analyzed spot forecasts as a “natural language” problem. Spot-forecast formats can vary by NWS WFO (Figs. 4 and 5). The table format provides time as a header with forecast values of elements that are requested valid at each time (Fig. 4). The non-table format provides the forecast of each requested element with a numeric value sometimes accompanied by a validation time (Fig. 5). Spot forecasts

occasionally include language that narrates the progression of MH, TWs, or HI values (Fig. 6). Some of these short narrations include specific times, and others use more ambiguous language, including “then,” “becoming,” “otherwise,” or “later,” creating a nebulous definition of time. Other forecasts include only one value or one range of values. The differences in the specificity of the forecast values for the requested BL elements create challenges in pattern recognition to isolate the appropriate value or range of values in the forecast and to build a database.

Text and numerical values associated with MH, TWs, or HI were extracted from spot forecasts that were issued and valid on the same day (e.g., the “today” forecast) within 50 km of an atmospheric sounding. The timing of the 0000 UTC atmospheric sounding provided guidance on selecting the appropriate forecast values from the corresponding spot forecasts. If only one numerical value or range of values existed, it was chosen as the verifiable forecast. However, because much of the data included text, measures were taken to select the latest forecast numerical value of the specified variables by interpreting key words and phrases such as “then,” “becoming,” “late in the afternoon,” and “early in the evening.” This ensures that the forecast numerical value chosen should coincide with the 0000 UTC sounding because the “today” portion of the spot forecast usually is valid until dark (sunset). If multiple numerical values existed with any of the “afternoon” or “evening” key words or phrases, the lowest and highest values were joined to form a forecast range. This method of parsing through spot forecasts and extracting values could lead to potential errors and biases. However, the method was consistently applied to all spot forecasts with many precautions—including noting very low or high values—to protect against obtaining erroneous or incorrect data.

Atmospheric soundings and their corresponding spot forecasts were removed if the data were missing for any of the necessary variables below 200 hPa. Conversely, spot forecasts missing pertinent information, such as elevation, variable units, or numerical forecast values, were removed along with their corresponding atmospheric soundings. Figs. 7–9 show the number of spot forecasts that exist within 50 km of an atmospheric sounding location by year, variable, and category and that remain before and after quality control measures. Of all available spot forecasts, 6.5% occurred within 50 km of an atmospheric sounding location, and 80%, 85%, and 89% of all MH, TWs, and HI spot forecasts, respectively, within the 50-km dis-

.TODAY...						
TIME (EDT)	6 AM	8 AM	10 AM	NOON	2 PM	4 PM
SKY.....	MCLDY	MCLDY	MCLDY	MCLDY	PCLDY	PCLDY
WEATHER TYPE....	RNSHWR	RNSHWR	RNSHWR	RNSHWR	RNSHWR	RNSHWR
CHC PRECIP (%)..	20	20	20	20	20	20
TEMP (F).....	84	84	86	87	88	87
RH (%).....	74	73	69	67	66	67
20 FT WIND.....	S 5	S 6	S 6	S 5	S 4	S 4
20 FT WIND GUST..	10	10	10	10	5	5
EYE LEVEL WIND..	S 5	S 6	S 6	S 5	S 4	S 4
EYE LVL WND GST..	10	10	10	10	5	5
MIX HGT (FT)....	700	1000	2700	4500	4500	4500
TRANSPORT WIND..	S 14	S 15	S 15	S 14	S 14	S 14
DISPERSION.....	12	14	18	47	44	44
LAL.....	2	2	2	2	2	2

Figure 4. Example of a spot forecast in a table produced by NWS WFO Key West, FL.

```
.TODAY...
SKY/WEATHER.....PARTLY CLOUDY(30-40%) UNTIL 1000...THEN MOSTLY
                    SUNNY UNTIL 1500...THEN PARTLY CLOUDY(50-60%).
MAX TEMPERATURE.....67-71.
MIN HUMIDITY.....16-23%.
20-FOOT WINDS.....WEST WINDS 5-13 MPH UNTIL 0900...THEN SOUTH
                    AROUND 6 MPH UNTIL 1200...THEN SOUTHEAST
                    5-10 MPH.
TRANSPORT WINDS.....SOUTHWEST 8-14 MPH.
MIXING HEIGHT.....BELOW 1000 FT AGL UNTIL 1000. 6400-7400 FT AGL
                    AFTER 1200.
SMOKE DISPERSAL.....POOR UNTIL 1000...THEN FAIR UNTIL 1200...THEN
                    GOOD UNTIL 1700...THEN FAIR.
```

Figure 5. Example of a spot forecast from NWS WFO Albuquerque, NM, that uses a mix of text and numerical values.

```
.TODAY...
SKY/WEATHER.....MOSTLY SUNNY (20-30 PERCENT).
                    TEMPERATURE.....42 AT IGNITION...MAX 60.
                    MIN HUMIDITY.....65 PERCENT AT IGNITION...MIN 25 PERCENT.
WIND (20 FT).....WINDS NORTH AT 1 MPH AT IGNITION...OTHERWISE
                    WEST WINDS 5 TO 6 MPH EARLY IN THE MORNING BECOMING
                    LIGHT.
CHANCE OF PCPN.....0 PERCENT.
LAL.....1.
HAINES INDEX.....3 OR VERY LOW POTENTIAL FOR LARGE PLUME DOMINATED
                    FIRE GROWTH.
MIXING HEIGHT.....2500 FT AGL AT IGNITION...INCREASING TO 2500-3500
                    FT AGL LATE IN THE MORNING...THEN INCREASING TO
                    4500-5500 FT AGL EARLY IN THE AFTERNOON INCREASING
                    TO 5500-6500 FT AGL LATE IN THE AFTERNOON DECREASING
                    TO 4000 FT AGL BY 1800 HOURS CST.
MIXING WINDS.....WEST 5 TO 6 MPH DECREASING TO 1 TO 3 MPH LATE IN THE
                    MORNING.
```

Figure 6. Example of a spot forecast with a more descriptive narrative using multiple forecasts valid at different times of the day and issued by NWS WFO Amarillo, TX.

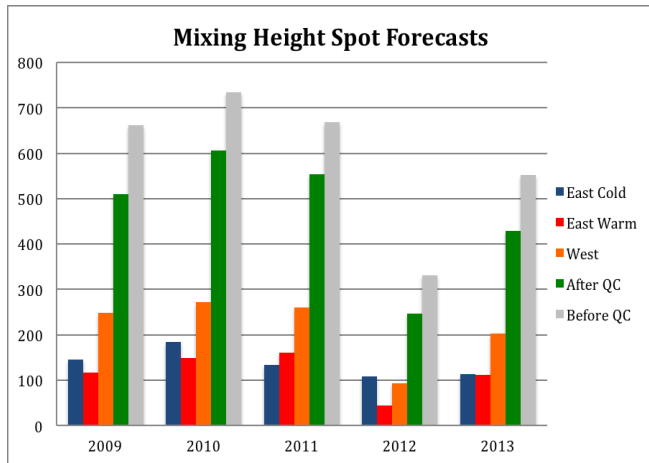


Figure 7. Number of spot forecasts with MH forecasts evaluated by year and category. Also shown is the number of spot forecasts before and after quality control measures.

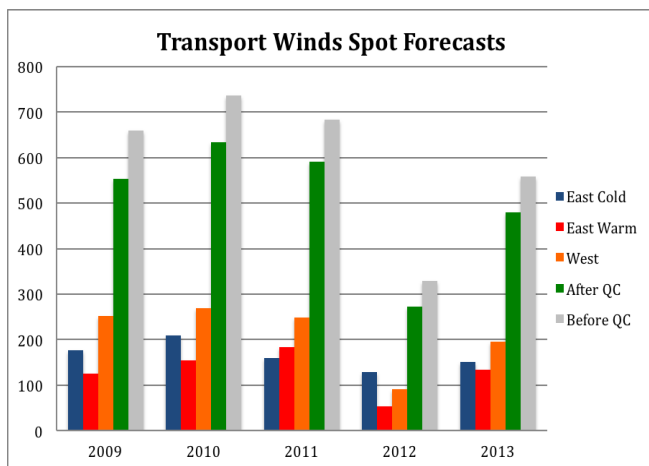


Figure 8. Same as Fig. 7 except for TWs.

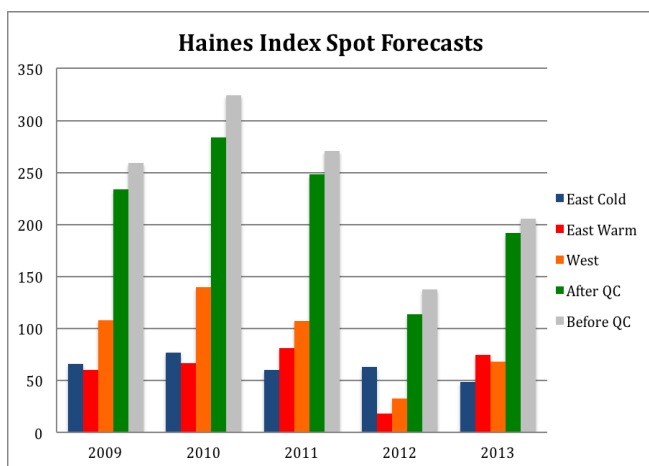


Figure 9. Same as Fig. 7 except for HI.

tance threshold were analyzed after the quality control measures.

d. Calculating mixing height, transport winds, and Haines index

The Holzworth, Stull, and Richardson methods of determining MH were calculated for the analysis. The Holzworth method defines MH as the height where Θ exceeds the surface Θ (Holzworth 1967). A rigorous application of the Stull method explores the entire vertical atmospheric profile to identify all areas of instability using Θ_v (Stull 1988, 1991). However, for verification analysis, MH was determined by finding the height where Θ_v exceeded the surface Θ_v (Fearon 2000; Fearon et al. 2015). The Richardson method involves calculating the bulk Richardson number at each level sampled within the atmospheric sounding until reaching a certain threshold delineating turbulent and laminar flow (Richardson 1920; Stull 2000). A traditional critical threshold of the bulk Richardson number is 0.25, but other BL studies have shown a value of 0.505, which was utilized for this verification as an appropriate threshold (Lee et al. 2009).

A temperature of 0.5°C was added to surface Θ and Θ_v to represent surface heating (Fearon 2000). The height closest to the elevation of the spot forecast within the atmospheric sounding was identified as the new surface. More than 95% of the analyzed spot forecasts resided at or above the elevations of the sounding sites. None of the remaining spot-forecast elevations were more than 350 m lower than the base elevation of the corresponding atmospheric sounding. If the spot forecast specified the MH as above ground level, then the height of that level minus the surface height produced the MH. If the spot forecast specified mean sea level (MSL), then the height of the identified level was used as the MH. This process was performed for each MH method and produced three MHs for each atmospheric sounding to be compared to the spot forecast.

TW speeds were calculated by averaging each atmospheric sounding level's wind speed at and below the MH. This calculation created three TW speeds for each sounding because of the three different methods of determining MH. Wind direction was not considered owing to the imprecise and nebulous language associated with wind direction in spot forecasts. Lammers and Horel (2014) cited similar reasoning for examining wind speed and not wind direction.

The HI is calculated following Haines (1988). The low-elevation (<200 m MSL) HI combines the temperature difference between 950 and 850 hPa and the dewpoint depression at 850 hPa. For mid-elevation (200–1000 m MSL), this changes the temperature

difference to between 850 and 700 hPa and the dewpoint depression to 850 hPa. For high-elevation sites (>1000 m MSL), HI combines the temperature difference between 700 and 500 hPa and dewpoint depression at 700 hPa (Haines 1988). These values are associated with coefficients that range from 1 to 3 and relate to the temperature difference and dewpoint depression (Table 3). The results from this calculation yield values from 2 to 6 with higher values representing a drier, more unstable lower atmosphere (Haines 1988).

Table 3. Temperature (T) and dewpoint (Td) differences and their reference values for all three elevations of the HI. Both reference values are added to calculate the HI (Haines 1988).

	Stability Term	Moisture Term
Low Elevation	950T–850T	850T–850Td
<200 m	1: <3°C	1: <5°C
	2: 4–7°C	2: 6–9°C
	3: >8°C	3: >10°C
Mid Elevation	850T–700T	850T–850Td
200–1000 m	1: <5°C	1: <5°C
	2: 6–10°C	2: 6–12°C
	3: >11°C	3: >13°C
High Elevation	700T–500T	700T–700Td
>1000 m	1: <17°C	1: <14°C
	2: 18–21°C	2: 15–20°C
	3: >22°C	3: >21°C
	Sum both terms to calculate the HI.	

e. Comparing spot forecasts and atmospheric sounding data

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 included a forecast range and an atmospheric sounding calculated variable occurred between the lower and upper bound of that forecast range, then the ME and MAE were 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 $\pm 5\%$ of the forecast value, which could have alleviated some of the bias toward range forecasts.

4. Analysis

a. Mixing height

Table 4 displays the mean forecast MH, the mean calculated MH using all three methods, and the number of spot forecasts that used one MH or a range of two MHs for all three regions. The EWS had the highest mean calculated Holzworth and Stull MHs; the highest mean forecast MHs were in the west; and the ECS had the lowest mean forecast and calculated MHs. WFOs in the west usually issued forecasts with one MH instead of a MH range (62.1%) compared to the ECS (48.3%) and EWS (35.6%). The Stull method consistently had the highest mean MH, and the Richardson method yielded the lowest mean MH in all three regions. The Holzworth method mean MH was in between the mean values of the Richardson and Stull methods, but its mean was closer to the Stull method's mean. The ECS had the largest ME for Stull and Holzworth methods, and the EWS had the largest ME using the Richardson method; all indicate overforecasting (positive MEs; Table 5). The west had the largest MAE for the Holzworth and Stull methods, but the EWS had the highest MAE for the Richardson method.

b. Transport wind speed

Table 6 displays the mean TW speed using each method, the mean TW speed forecast, and the number of forecasts with one TW speed or two TW speeds (e.g., 5.1–7.5 m s⁻¹). The forecasters in the west issued more TW speed forecasts with one just wind speed (40.7%) than ECS (19.4%) and EWS (18.3%), with all of these percentages lower than their respective MH percentages. The mean TW speed forecasts for all three regions are very close (<1 m s⁻¹). The west also has the highest mean wind speeds for each method, with the ECS having the lowest for each method. The EWS has the lowest MAE and ME for all three methods (Table 7). The west has a higher ME than EWS and has the largest MAE for all three methods. The ECS has the highest ME for all three methods. The west, ECS, and EWS all demonstrate an overforecasting bias (positive ME), with the strongest bias associated with the ECS.

c. Haines index

Table 8 shows the results for the low-, mid-, and high-elevation HI calculations. The HI is requested the least among the three variables at a rate about half of

Table 4. Mean MH for spot forecasts (using one forecast value or a range of values) and for each method. The number of spot forecasts utilizing one forecast value or a range is totaled for MH.

	Mean Forecast: 1 MH	Mean Forecast: MH Range	Mean Holzworth MH	Mean Stull MH	Mean Richardson MH	Number of 1 MH Forecasts	Number of MH Range Forecasts
West	1938 m	1466–1822 m	843 m	885 m	703 m	666	407
East Cold Season (ECS)	1287 m	789–1144 m	557 m	603 m	465 m	331	355
East Warm Season (EWS)	1578 m	1319–1687 m	1023 m	1107 m	623 m	207	374

Table 5. Mean errors (MEs) and mean absolute errors (MAEs) of mixing height (MH) for each method.

	ME (Spot–Holzworth)	ME (Spot–Stull)	ME (Spot–Richardson)	MAE (Spot versus Holzworth)	MAE (Spot versus Stull)	MAE (Spot versus Richardson)
West	370 m	341 m	562 m	618 m	618 m	659 m
East Cold Season (ECS)	465 m	426 m	549 m	553 m	535 m	583 m
East Warm Season (EWS)	406 m	340 m	752 m	529 m	503 m	786 m

Table 6. Same as Table 4 except for TW speed.

	Mean Forecast: 1 TW Speed	Mean Forecast: TW Speed Range	Mean Holzworth TW Speed	Mean Stull TW Speed	Mean Richardson TW Speed	Number of 1 TW Speed Forecasts	Number of TW Speed Range Forecasts
West	7.18 ms ⁻¹	5.07–7.51 ms ⁻¹	6.06 ms ⁻¹	6.18 ms ⁻¹	5.56 ms ⁻¹	424	618
East Cold Season (ECS)	6.18 ms ⁻¹	5.37–7.52 ms ⁻¹	4.33 ms ⁻¹	4.51 ms ⁻¹	4.75 ms ⁻¹	150	623
East Warm Season (EWS)	6.22 ms ⁻¹	4.95–7.09 ms ⁻¹	5.48 ms ⁻¹	5.64 ms ⁻¹	5.20 ms ⁻¹	97	529

Table 7. Same as Table 5 except for TW speed.

	ME (Spot–Holzworth)	ME (Spot–Stull)	ME (Spot–Richardson)	MAE (Spot versus Holzworth)	MAE (Spot versus Stull)	MAE (Spot versus Richardson)
West	0.24 ms ⁻¹	0.16 ms ⁻¹	0.64 ms ⁻¹	2.78 ms ⁻¹	2.52 ms ⁻¹	2.54 ms ⁻¹
East Cold Season (ECS)	1.36 ms ⁻¹	1.13 ms ⁻¹	0.93 ms ⁻¹	2.36 ms ⁻¹	2.18 ms ⁻¹	1.92 ms ⁻¹
East Warm Season (EWS)	0.04 ms ⁻¹	-0.11 ms ⁻¹	0.32 ms ⁻¹	1.88 ms ⁻¹	1.76 ms ⁻¹	1.62 ms ⁻¹

Table 8. MEs and MAEs of HI and mean HI for spot forecasts (using one forecast value or a range of values) and the total number of each type of forecast.

	ME HI	MAE HI	Mean HI	Mean of 1 HI Forecast	Mean of HI Range Forecast	Number of 1 HI Forecasts	Number of HI Range Forecasts
West	-0.89	1.07	4.22	3.33	3.0–4.0	455	1
East Cold Season (ECS)	-0.10	0.78	4.33	4.29	3.40–4.61	178	137
East Warm Season (EWS)	-0.20	0.47	4.27	4.03	3.20–4.54	196	105

MH and TWs. Most HI forecasts include one value, especially in the west, where all but one forecast had one value. The ECS has the highest mean HI, although it is surprising that the cold season would be associated with the highest HI or an unstable, dry lower atmosphere. The west has the largest MAE and ME, with ME indicating underforecasting (negative ME) by nearly one category (Table 8). The ECS exhibits the least amount of bias with the lowest ME, and the EWS has the lowest MAE.

5. Discussion

Of all three BL variables examined, MH forecasts performed the worst, and the west, ECS, and EWS all exhibited overforecasting (positive ME) of MH. All three methods of forecast MHs exhibited MEs of at least 30% of their respective mean MHs with most of the MH MEs >50%. Additionally, the MAEs of all three MH methods were at least 45% of their respective MH means, with several near or larger than their means (Tables 4 and 5). The large magnitude of the errors was a surprise, but verifying MH forecasts is difficult for these reasons: 1) multiple definitions of MH; 2) multiple methods of calculating MH and no information on what approach forecasters are most likely to use; and 3) the potentially large distance and terrain variability (as much as 50 km from atmospheric sounding locations to the spot-forecast locations). Fearon et al. (2015) showed that Holzworth, Stull, and Richardson methods underestimate MH, which also could help explain the overforecasting and contribute to the large MAEs. TW speeds usually exhibited smaller errors relative to their means than MH (Table 7), which could be due to averaging the wind speeds through the calculated ML. Additionally, NWS forecasters use a range of values more often in TW speed forecasts than in MH forecasts. HI is less sensitive to the issues raised above, and thus performed nearly as well or better by most measures than TWs and MH (Table 8).

The inconsistency and ambiguity of the language used in many of the spot forecasts among the NWS WFOs creates challenges for verification and leaves the accuracy of the spot forecasts vulnerable to different interpretations. The verification process treats equally a spot forecast with one MH and a spot forecast that describes multiple MHs during the course of a day. Certain methods implemented could create an accuracy bias toward forecasts that provided a range of

values or multiple values valid at different times during the day.

Any forecast verification must define accuracy. The spot-forecast user requires accuracy for the requested variables, but the precision of those variables remains dependent on the particular needs of the user (OFCM 2007, 2011). Penalizing a spot forecast with one value or one value valid in the afternoon by not including a certain level of acceptable inaccuracy, such as $\pm 5\%$ of the total forecast, may not be appropriate. However, no current standards exist for determining acceptable error in spot forecasts or specifying what the contents of each spot forecast variable should include, as noted in OFCM (2007, 2011). The Joint Fire Science Program (JFSP) recently has funded projects to examine weather data in the context of decision-making, the research outcomes of which could be relevant to improve spot forecasts.

6. Summary and recommendations

The results of this study demonstrate that TWs and HI spot forecasts exhibit relatively small MEs and MAEs compared to the relatively large MH spot forecasts errors. The MH result in particular underlines issues in MH forecast consistency and methods and subsequent limitations for verification. Although these results are informative for understanding the current state of MH, TWs, and HI in spot forecasts and for potential improvements of forecasts and verification methods, they are not absolute. For example, Fearon et al. (2015) highlighted the challenges of MH calculation and forecasting and, hence, the verification of this type of forecast. These challenges include the spatial and temporal representativeness of the atmospheric soundings to a specific location and validation time and using independent sources (e.g., satellite optical depth data) to generate MH. With atmospheric soundings spread across four time zones, at various latitudes, and occurring throughout the year, 0000 UTC atmospheric sounding data may not be representative of the day's MH, TWs, or HI because of the increasing influence of the nocturnal BL. Additionally, complex terrain causes significantly different BL conditions over short distances. One approach to address the representativeness of 0000 UTC atmospheric soundings would be to use vertical profiles at nearby grid points from operational numerical model analyses (e.g., North American Mesoscale Model, Rapid Refresh, or High-Resolution Rapid Refresh; rapidrefresh.noaa.gov). However, using such a model-based verification ap-

proach could focus on the degree to which the forecaster deviated from model output as these models are used operationally.

The OFCM (2007, 2011) and NOAA Science Advisory Board (2008) made similar recommendations regarding accuracy requirements, improvements, and verification statistics for both surface and upper-air elements in spot forecasts. Lammers and Horel (2014) and this paper represent the beginning of a process to address these recommendations and demonstrate that spot-forecast verification necessitates a more nuanced approach than just aggregating statistics. Echoing Lammers and Horel's (2014) principal recommendation, forecasters and end users should develop a framework that allows flexibility in deciding how and what to verify from spot forecasts. Additionally, the results from this spot-forecast verification highlight some appropriate recommendations for any future operational attempt to evaluate spot forecasts, such as that planned by the NWS Performance Branch:

- 1) The consistency of the information provided by the spot forecasts needs to be improved. Lammers and Horel (2014) echo this sentiment with the recommendation to "isolate quantitative numerical values separately from qualitative alphabetical descriptors." With the understanding that different regions have different climates and user needs, the forecast values and the description of these values for each variable requested still need standardization. Lammers and Horel (2014) discussed the importance of the qualitative information in spot forecasts for fire managers, but also recommended extracting basic forecast numerical information from the spot forecast for verification purposes.
- 2) 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. Developing a framework for spot-forecast verification allows forecasters to aggregate data and evaluate spot forecasts quickly. Lammers and Horel (2014) endorsed this recommendation and noted that forecasters evaluating spot forecasts with local knowledge would be an improvement over "depending on bulk statistical metrics accumulated on national scales." More robust tools also are needed to evaluate spot forecasts, including those in areas of com-

plex terrain and not near weather stations or sounding locations.

- 3) 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 provides insight into possible sources of consistent errors that may lead to improving forecasts. Consistent sources of errors could include utilizing only one MH method or recognizing the limitations of forecasts in complex terrain. Understanding these sources of errors could help standardize specific methods for determining which lower-atmospheric variables are more useful in different regions or during different seasons.
- 4) 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.

Acknowledgments. We thank Matt Lammers for providing the NWS spot forecasts (via Virgil Middendorf, NWS Billings WFO), sharing code, and providing guidance throughout the research. We also thank Drs. Michael L. Kaplan, John F. Mejia, and Peter J. Weisberg for reviewing the manuscript before submission and Mathew G. Fearon for his expertise in Python and mixing-height determination. Funding of this project was provided by the JFSP under grant 12-1-05-03.

REFERENCES

- AirFire, cited 2015: Glossary. United States Department of Agriculture, Forest Service. [Available online at www.airfire.org/projects/wfdss-aq/help/glossary/.]
- Beyrich, F., 1997: Mixing height estimation from sodar data—A critical discussion. *Atmos. Environ.*, **31**, 3941–3953, [CrossRef](#).
- Brier, G. W., and R. A. Allen, 1951: Verification of weather forecasts. *Compendium of Meteorology*, T. F. Malone, Ed., Amer. Meteor. Soc., 841–848.
- Brown, B. G., and A. H. Murphy, 1987: Quantification of uncertainty in fire-weather forecasts: Some results of operational and experimental forecasting programs. *Wea. Forecasting*, **2**, 190–205, [CrossRef](#).
- Fearon, M. G., 2000: The use of nonlocal static stability to determine mixing height from NCEP Eta model output over the western U.S. M.S. thesis, Dept. of

- Atmospheric Sciences, University of Nevada, 161 pp. [Available online at www.cefa.dri.edu/Publications/mfearon_msthesis.pdf.]
- _____, T. J. Brown, and G. M. Curcio, 2015: Establishing a national standard for methodology for operational mixing height determination. *J. Operational Meteor.*, **3** (15), 172–189, [CrossRef](#).
- Haines, D. A. 1988: A lower atmosphere severity index for wildlife [sic] fires. *Natl. Wea. Dig.*, **13** (2), 23–27. [Available online at www.nwas.org/digest/papers/1988/Vol13-Issue2-May1988/Pg23-Haines.pdf.]
- Holzworth, G. C., 1967: Mixing depths, wind speeds and air pollution potential for selected locations in the United States. *J. Appl. Meteor.*, **6**, 1039–1044, [CrossRef](#).
- Joliffe, I. T., and D. B. Stephenson, Eds., 2003: *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. Wiley & Sons, 240 pp.
- Lammers, M. R., and J. D. Horel, 2014: Verification of National Weather Service spot forecasts using surface observations. *J. Operational Meteor.*, **2** (20), 246–264, [CrossRef](#).
- Lee, P., and Coauthors, 2009: Impact of consistent boundary layer mixing approaches between NAM and CMAQ. *Environ. Fluid Mech.*, **9**, 23–42, [CrossRef](#).
- Miller, M. E., 1967: Forecasting afternoon mixing depths and transport wind speeds. *Mon. Wea. Rev.*, **95**, 35–44, [CrossRef](#).
- NWS, cited 2015: Transport Wind (kts) (definition). [Available online at graphical.weather.gov/definitions/defineTransWind.html.]
- NOAA Science Advisory Board, 2008: Fire weather research: A burning agenda for NOAA. NOAA, 98 pp. [Available online at www.sab.noaa.gov/Reports/2008/FWRWGreportFINALfromSABtoNOAA_11_03_08.pdf.]
- OFCM, 2007: National wildland fire weather: A summary of user needs and issues. FCM-R33-2007. Federal Coordinator for Meteorological Services and Supporting Research, Silver Spring, MD, 65 pp. [Available online at www.ofcm.gov/r33-r34/r33-3007/Wildland%20Fire%20Weather%20User%20Needs%20110517%20FINAL.pdf.]
- _____, 2011: Wildland fire weather: Multi-agency portfolio of current and in-development capabilities to support user needs. Federal Coordinator for Meteorological Services and Supporting Research, Silver Spring, MD, 212 pp. [Available online at www.ofcm.gov/r33-r34/r34-2011/R34.pdf.]
- Potter, B. E., J. A. Winkler, D. F. Wilhelm, R. P. Shadbolt, and X. Bian, 2008: Computing the low-elevation variant of the Haines index for fire weather forecasts. *Wea. Forecasting*, **23**, 159–167, [CrossRef](#).
- Richardson, L. F., 1920: The supply of energy from and to atmospheric eddies. *Proc. Roy. Soc. Lond. A*, **97**, 354–373, [CrossRef](#).
- Stull, R. B., 1988: *An Introduction to Boundary Layer Meteorology*. Kluwer Academic Publishers, 666 pp.
- _____, 1991: Static stability—An update. *Bull. Amer. Meteor.*, **72**, 1521–1529, [CrossRef](#).
- _____, 2000: *Meteorology for Scientists and Engineers*. 2nd ed. Cengage Learning, Inc./Brooks Cole, 502 pp.
- Wallace, J. M., and P. V. Hobbs, 2006. *Atmospheric Science: An Introductory Survey*. 2nd ed. Elsevier/Academic Press. 504 pp.
- Werth, P., and R. Ochoa, 1993: The evaluation of Idaho wildfire growth using the Haines index. *Wea. Forecasting*, **8**, 223–234, [CrossRef](#).
- _____, and Coauthors, 2011: Synthesis of knowledge of extreme fire behavior: Volume 1 for fire managers. U. S. Department of Agriculture, Forest Service, Pacific Northwest Research Station, Portland, Oregon. General Technical Report PNW-GTR-854. 158 pp. [Available online at www.fs.fed.us/pnw/pubs/pnw_gtr854.pdf.]
- Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. 2nd ed. Elsevier/Academic Press, 648 pp.