Verification of Red Flag Warnings Across the Northwestern U.S. as Forecasts of Large Fire Occurrence

A Thesis
Presented in Partial Fulfillment of the Requirements for the
Degree of Master of Science
with a
Major in Natural Resources
in the
College of Graduate Studies
University of Idaho
by
Joshua M. Clark

Major Professor: Crystal Kolden, Ph.D.
Committee Members: John Abatzoglou, Ph.D.; Alistair Smith, Ph.D.
Department Administrator: Charles Goebel, Ph.D.

May 2020

Authorization to Submit Thesis

This thesis of Joshua M. Clark, submitted for the degree of Master of Science with a Major in Natural Resources and titled "Verification of Red Flag Warnings Across the Northwestern U.S. as Forecasts of Large Fire Occurrence," has been reviewed in final form. Permission, as indicated by the signatures and dates below, is now granted to submit final copies to the College of Graduate Studies for approval.

Major Professor: ____ Date: ____________

Crystal Kolden, Ph.D.

Committee Members: ____ Date: ____________

John Abatzoglou, Ph.D.

____ Date: ____________

Alistair Smith, Ph.D.

Department

Administrator: ____ Date: ____________

Charles Goebel, Ph.D.

Abstract

Red Flag Warnings (RFWs) issued by the National Weather Service in the United States (U.S.) are an important fire early warning system based on forecasts of critical fire weather that foster fire activity including the occurrence of large fires. However, verification of RFWs as they relate to fire activity is lacking, thereby limiting means to improve forecasts as well as increase value for end-users. We evaluated the efficacy of RFWs as forecasts of large fire occurrence for the Northwestern U.S and found favorable performance broadly across the area, along with substantial skill and improvement over reference forecasts. We further demonstrate that the skill of RFWs is significantly higher for lightning-ignited large fires and for forecasts issued during periods of high fuel dryness. The results of this first verification study of RFWs lay the groundwork for future efforts towards improving the relevance and usefulness of RFWs to better serve the fire community and public.

Acknowledgements

First, I wish to thank my advisor, Dr. John Abatzoglou, for his incredible patience, kindness, and advice regarding this research and more broadly, my aims as a scientist. John guided and reassured me through countless iterations on this work, and his brilliance, humor, and enthusiasm made him the most ideal mentor I could ask for. I also wish to thank the members of my committee and my colleague Nick Nauslar, who provided much feedback on this paper and greatly assisted with answering any questions I had.

My co-workers, Kirk Davis and Sarah Krock, are owed considerable acknowledgement for their assistance in reviewing this research, devising GIS analysis techniques for processing the data, and for *constantly* encouraging me to complete this paper. My supervisors, Karen Arnold and Angie Lane, unquestionably supported this research as part of my scientific role and for career development — it would have likely been much more difficult to complete this without their support.

I wish to also acknowledge my father, who inspired in me a love for weather some twenty years ago. Although I am sure he would rather I take up severe storm forecasting, I know he would love fire weather just the same if only there were big fires back in Kentucky.

Lastly, I want to thank my wife, Jenna, who has supported me since I first headed down this path. She is the ultimate partner, friend, and counselor, and mother to our two cats (and one dog on the way!). I am positive I would not have accomplished a thesis without her encouragement.

Table of Contents

List of Tables

List of Figures

1. Introduction

Wildland fire plays an important role as a natural and increasingly anthropogenic disturbance found in most vegetated ecosystems globally (Bowman *et al.*, 2009) and generally serves to promote healthy, resilient landscapes. However, when fire threatens to directly impact human life and property, it may be considered hazardous (Moritz *et al.*, 2014), prompting land management action to avoid potentially catastrophic consequences. In recent decades, longer periods of critical fire weather (Jolly *et al.*, 2015) juxtaposed by human-caused fire activity (Balch *et al.*, 2017) have expanded the threat of hazardous fires in the United States (U.S.), making land management objectives more difficult and costlier to achieve while also placing the safety of fire suppression personnel and the public at greater risk. These effects have been demonstrated by recent hazardous fire events in the U.S. (Brewer and Clements, 2020; Nauslar *et al.*, 2018; Balch *et al.*, 2018), and are expected to continue or worsen due to anthropogenic climate change (Barbero *et al.*, 2015; Moritz *et al.*, 2014) and the growth of wildland-urban interfaces (Radeloff *et al.,* 2018).

Many warning systems have been developed for fire hazards (de Groot *et al.*, 2014). Comprehensive fire danger systems integrate weather, fuels, and climate information to generate daily predictions regarding the fire environment and fire behavior characteristics (Deeming *et al.*, 1983; Van Wagner, 1987). These outputs are generally considered in fire management plans as actionable criteria that prompt some prevention, preparedness, or resource allocation decision (de Groot *et al.*, 2014) and have seen widespread adoption in the wildland fire community. Novel systems, such as the Severe Fire Danger Index and Santa Ana Wind Threat Index, leverage advancements in gridded meteorological forecast data along with fuels and fire behavior information to provide local-to-regional predictions of fire danger (Jolly *et al.*, 2019; Rolinski *et al.*, 2016), while other systems predict fire danger directly from fire weather conditions independent of fuels (Srock *et al.*, 2018, Erickson *et al.*, 2016). While these systems serve as important tools for understanding hazardous fire risk, many are generated independent of real-time human input that may be important for deciphering rapidly changing fire environment factors that drive the most extreme fire activity.

In the U.S., Red Flag Warnings (RFWs) issued by meteorologists at the National Oceanic and Atmospheric Administration (NOAA) National Weather Service (NWS), are forecasts of critical fire weather conditions, that, in the presence of sufficiently dry fuels, could lead to abundant new ignitions and/or rapid spread and growth on existing fires (Figure 1.1) (NWS, 2017). Forecasters at local weather forecast offices (WFOs) issue RFWs on short timescales (imminent to 48-hr lead times) and consider numerous quantitative and qualitative meteorological parameters (e.g. relative humidity, wind speed, atmospheric stability, dry frontal passage, lightning) along with some measure of fuel dryness. These parameters are considered as criteria for RFW issuance (Table 1.2) based on the expert opinion and analyses at WFOs and vary geographically by sub-regional fire weather zones (FWZs). Although this arrangement allows for greater flexibility to accommodate local conditions most conducive to area fire initiation and growth, it has resulted in many different criteria across FWZs, complicating the messaging intent of these forecasts. Further, these numerous criteria have limited efforts to verify the performance of RFWs, and no peer-reviewed studies could be located that would otherwise give information on possible refinements to these forecasts.

RFWs are similar to other NWS early warning products for natural hazards (e.g. tornadoes, severe thunderstorms, flash floods) as they are issued for high-impact events that threaten life and property. Many verification studies have been conducted on the performance of these other hazard forecasts to increase the quality of the system and value to end-users (Brooks and Correia, 2018; Obermeier and Anderson, 2014; Clark *et al.*, 2014; McGovern *et al.*, 2014). The NWS maintains internal verification statistics for RFWs, but these are computed for weather that could *lead to* hazardous fires, not actually the occurrence of hazardous fires (unlike warnings for tornadoes or flash floods). It is useful, however, to explore the performance of RFWs as direct forecasts of hazardous fire activity to assess the ability of RFWs to discriminate those fires which have the greatest effect on resources and human values, and thus are of most concern to end-users.

Here, we consider occurrences of new large fires (LFs) as hazardous fires and treat RFWs as dichotomous forecasts of these fire occurrences across subregions of the Northwestern U.S.. Since LF occurrence is complicated by many non-meteorological factors, we examine how verification of RFW performance varies across different LF size thresholds, fire causes, and land cover. Finally, we assess how forecast performance varies as a function of relative fuel dryness, which while being a criterion considered in many RFWs, is applied heterogeneously across zones. This study provides a first known effort to evaluate the added value of RFWs for actualized LF activity that is of key importance for fire suppression. Results of this study may help to both refine RFWs and identify reasons why forecast performance and skill varies across different regions and LF characteristics.

2. Materials and Methods

2.1 Study Location

The study area is composed of the county warning areas of eight NWS WFOs of the northwest contiguous U.S.: Seattle (SEW), Spokane (OTX), Missoula (MSO), Pocatello (PIH), Boise (BOI), Medford (MFR), Pendleton (PDT), and Portland (PQR) (Figure 2.1). This covers a diverse set of regions we herein refer to as the Northwest, including the coastal areas of Washington and Oregon, the Cascade Mountains and Blue Mountains, much of the Northern Rockies, the Columbia Basin, and portions of the Great Basin. Fire season in the Northwest is generally shorter and more well-defined compared to other fire-prone regions across the U.S. (Werth *et al.*, 2016), with ~86.1% of all wildfires and ~99.7% of total burned area occurring in the months of June - October during 2006 - 2015. Across the region, the number of lightning ignited fires and human ignited fires is nearly the same, although lightning fires comprise ~83.6% of total burned area while human ignited fires account for only ~16.42% of total burned area (Short, 2017). Smaller, generally human ignited fires are more commonly found west of the Cascade Mountains due to less lightning frequency (Agee, 1993; Abatzoglou *et al.,* 2016) and a larger wildland-urban interface component (Balch *et al.*, 2017).

2.2 Datasets

Spatial data of RFWs were obtained from the Iowa State University Iowa Environmental Mesonet (IEM) archive of NWS watches and warnings (IEM, 2019). This dataset includes the WFO, FWZ, issuance date/time, and expiration date/time with each warning. We simplified RFWs that span multiple days into records for each day such that if a RFW was active for any portion of a calendar day it was considered a RFW day for that FWZ. The resulting dataset contains 8,940 RFW forecast days during 2006–2015.

Point-location fire records for the same 10-year period were obtained from the Fire Program Analysis fire occurrence database (Short, 2017). This data includes fire discovery date, final fire size, and fire cause. Records were reduced to the study area resulting in 64,122 fires that were assigned to the corresponding FWZ that existed at the time of fire discovery date. This latter point is important as FWZ boundaries have changed throughout the entire period due to anticipated improvements in the quality of forecasts and ability to adequately warn affected populations (NWS, 2018). We designate LFs for each FWZ as the largest 10% of fires within each zone per Nagy *et al.* (2018), but also assess the sensitivity of results to choice of fire size by considering the largest 5% and 20% of fires. The resulting number of LFs across the entire study region were 10,402, 5,341, and 2,775 for when fire sizes were >= 80th, 90th, 95th percentile sizes, respectively. For the purpose of tabulating LFs, we assumed that fires reached size thresholds on the date of discovery as detailed fire progression information was unavailable.

We additionally explore how RFW performance varies by fire cause, land cover, and a measure of relative fuel dryness. These subsequent tests are constrained to LFs above the 90th percentile size threshold. To more effectively determine performance for just human- and lightning-caused fires, we eliminated those which had an 'unknown' cause, leaving 2,309 lightning-caused and 2,890 human-caused LFs. However, we retained these 'unknown' caused fires in the other tests to maximize the frequency of events available for computing performance statistics. We then classified LFs as burning primarily in forested lands and those primarily burning in non-forested lands using the 0.5 km MODIS-based global land cover climatology dataset (Broxton *et al.*, 2014). Due to the absence of fire perimeters, we approximated perimeters by assuming a circular fire with an area equivalent to the final fire size. Fires which had >50% pixels classified as forested were assigned as such and those <50% were assigned as non-forested vegetation. A total of 3,073 LFs were classified as forested while the remaining 2,466 were considered non-forested.

Lastly, energy release component (ERC) percentile values were assigned to each fire using co-located ~4-km gridded data from gridMET (Abatzoglou, 2013) to represent fuel dryness. These percentiles were calculated by pooling data for the entire calendar year during the 2006–2015 period. ERC is defined as the total available energy within the flaming front of a fire calculated from the U.S. National Fire Danger Rating System, here using a commonly used fuel model (dense conifer). ERC is often used in fire business decision-making as it is a good measure of cumulative fire danger as it gives higher weighting to heavier fuel types that tend to reflect seasonal drying trends (Freeborn *et al.*, 2015), yet does not consider the short-term effects of wind in daily fluctuations of fire weather. Several studies have shown a strong relationship between interannual fire activity and ERC (Abatzoglou and Kolden, 2013; Barbero *et al.*, 2014). Most commonly, the 90th percentile ERC threshold is used to represent high fire danger across particular geographic regions (Heinsch *et al.*, 2009; Dalton *et al.*, 2015, Jolly *et al.*, 2019), and we adopt this threshold in our analysis. A total of 2,160 LFs occurred with ERC >= 90th percentile while 3,307 occurred when ERC < 90th percentile.

2.3 Analysis Methods

To quantify forecast performance, we treated RFWs as dichotomous forecasts for new LFs and constructed a contingency table that shows the frequency of forecasts and LF occurrences. The contingency table is a standard forecast verification tool used to compute performance statistics of nonprobabilistic forecasts of discrete predictands (Murphy and Winkler, 1987; Doswell *et al.*, 1990; Wilks, 2006) and has the general form:

		Observed Event?	
		Yes	**No**
Forecast Issued?	**Yes**	Correct Positive (a)	False Alarm (b)
	No	Miss (c)	Correct Negative (d)*

** Not calculated*

Here, correct positive forecasts (i.e., *hits*) are RFWs where a new LF occurred on or within one day following the forecast in the same FWZ. This 2-day period was used to accommodate typical delays in fire reporting and is consistent with other studies that consider meteorological conditions immediate to fire occurrence using LF databases (Abatzoglou *et al.*, 2018). If a RFW was issued but no LFs were observed coincident with or one day after the forecast date, the forecast is classified as a *false alarm*. A forecast is classified as a *miss* when there was no forecast issued for a FWZ during the 2-day period but a LF occurred. Since RFWs are forecasts of rare events (similar to tornadoes, flash floods) calculating correct negative forecasts are of little value due to the overwhelming amount of days where no event was observed or forecast (e.g., Clark *et al.*, 2014; Obermeier and Anderson, 2014; Vaughan *et al.*, 2017). Thus, we omit correct negatives in our analysis (Mason, 2003). An example of this classification scheme for a particular day is given in Figure 2.2. On 10 August 2015, 32 RFWs (shown as hits or false alarms) were issued and 90th percentile sized LFs which occurred on 10–11 August were used to classify forecasts. The resulting number of hits, misses, and false alarms for this RFW day were 11, 11, and 21, respectively.

From the contingency table, we computed the following performance measures: probability of detection (POD), false alarm ratio (FAR), and critical success index (CSI). The relevant equations and definitions of these metrics are listed in Table 2.1. These measures are frequently reported together in the case of rare event forecasts because of a shared lack of consideration given to correct negatives (Schaefer, 1990). Further, if we consider the formulae in Table 2.1, we can rearrange terms to show that CSI is a nonlinear function of POD and FAR given by:

$$CSI = [(POD)^{-1} + (1 - FAR)^{-1} - 1]^{-1}$$

For rare events such as LFs, it is important to note that values will seldom approach CSI = 1 (and may instead be much closer to CSI = 0) due to the decreased frequency of events and increased frequency of times where no event was forecast or observed.

Most commonly, forecast systems are evaluated against persistence forecasts, random forecasts, or climatological forecasts by determining the 'skill score' of the compared forecasts for a particular performance metric (Wilks, 2006). These skill scores provide an analytic guide to measure the added value of a forecast relative to a reference forecast. In its most generic form, the skill score with respect to some performance metric (M) is defined as:

$$M_{ss} = \frac{M_{FORECAST} - M_{REFERENCE}}{M_{PERFECT} - M_{REFERENCE}}$$

where $M_{FORECAST}$ is the performance metric (such as POD or CSI) of the forecast system being evaluated, $M_{REFERENCE}$ is the same metric but for the reference forecast, and $M_{PERFECT}$ is the scalar value of a perfect forecast for that particular metric (for example, $POD_{PERFECT} = 1$ if POD was being evaluated). If $M_{ss} > 0$, improvement over the reference forecast can be inferred.

Since RFWs and LF occurrence within the study area are generally confined to the summer months, reference forecasts need to preserve this seasonality for fair comparisons.

Climatological and persistence forecasts are commonly used as reference forecasts for skill score calculations, particularly for continuous variables (Potts, 2003; Mason, 2004; Warner, 2011). For our case, a persistence-based reference would not be suitable given the serial correlation of these forecasts from one day to the next that would make it difficult to separate skill of the actual forecasts from the reference (Mittermaier, 2008). A randomly generated forecast provides a reference completely independent of the observations (a truly no-skill forecast) but sacrifices the seasonal relationship we are trying to preserve. Thus, we define a 'random climatology' forecast to achieve an independent set of reference forecasts that retain a similar seasonal distribution to the actual RFWs. To achieve this, a resampling procedure was implemented where actual RFW days are resampled for any date within ± 15 days among all years of the study period resulting in a reference set of the same size. These reference forecasts were then assessed against the observations using the performance metrics and skill scores were computed against the actual forecasts. This process was repeated one hundred times to obtain a robust sample of statistics and skill scores and median values were selected for presentation in the results. To test the statistical significance of RFW forecast skill versus the reference, we reject the null hypothesis of RFW skill being the same as the reference forecast if 95% of the sample skill scores are greater than zero.

We also report results in terms of relative change between the RFWs and reference as another measure of skill since skill scores may be unintuitive to fire managers and other readers external to the weather forecasting community. This formula is given as:

$$Rel.Change\ (\%) = \frac{M_{FORECAST} - M_{REFERENCE}}{M_{REFERENCE}} \times 100$$

In addition to reporting performance metrics and skill for each WFO, we compute area-wide statistics by aggregating the number of hits, misses, and false alarms from each WFO into one contingency table representing all eight offices.

3. Results

3.1 Performance as forecasts of LF occurrence

We found that RFWs exhibited skill as forecasts of LF occurrence across WFOs in the Northwestern US (Figure 3.1). Area-wide POD was 0.29, while POD_{SS} was 0.18 and POD relative improvement was 124.1% over the reference forecast. Six of the eight WFOs demonstrated >100% POD relative improvements. POD was the lowest for two offices covering the populated and mesic portions of the study area (SEW and PQR), although relative improvements for capturing LF occurrence in these areas were high.

3.2 Fire size

Area-wide POD increased with larger fire size thresholds, although this was countered by increased FAR values due to decreasing event frequency (Table 3.1). Similarly, area-wide skill scores showed improved POD_{SS} from 0.13 for 80th percentile sized LFs to 0.23 for 95th percentile sized LFs. The area-wide relative improvement of POD increased from 99.4% to 138.2% as size thresholds increased. Differences in POD as a function of fire size were variable among the WFOs; across all fire size thresholds, the highest performance was shown for regions with the largest fire sizes (BOI, PIH, and PDT) while notably lower performance was observed for regions west of the Cascades Mountains (PQR and SEW). Five WFOs demonstrated relative improvements over the reference forecast > 100% regardless of fire size threshold.

3.3 Fire cause

Area-wide POD, POD_{SS}, and relative improvement of POD were notably higher for lightning-caused LFs than for human-caused LFs (Figure 3.2). For example, the area-wide POD was 0.46 for lightning-caused LFs and 0.17 for human-caused LFs. Further, POD_{SS} was 0.34 for lightning-caused LFs and only 0.08 for human-caused LFs. While skill scores for human-caused LFs were low, they showed substantial relative improvement (~78.4%) above the reference.

Similar to area-wide results, we generally found higher skill for lightning-caused fires than human-caused fires across WFOs. Figure 3.2b shows this as points for each WFO where higher POD and lower FAR pairs result in increased CSI values. WFOs with the fewest lightning-caused LFs (PIH, PQR, and SEW) showed the lowest POD and highest FAR. Although more human-caused LFs occurred across the region, FAR values for these fires were higher than lightning-caused LFs for five of the eight WFOs. Skill scores demonstrated improvement for both cause types across regions, although the improvement among human-caused fires was low.

3.4 Land cover type

We found little discernible difference in performance and skill scores calculated for LFs stratified by forest and non-forest land cover (Table 3.2). Area-wide POD was slightly higher for non-forested fires but FAR and CSI were nearly the same. Both land cover types showed area-wide relative improvement >125% above the reference. We found statistically significant skill for all RFWs by land cover except for non-forested LFs in SEW.

3.5 Fire danger

RFW forecasts conditioned on being coincident with high fire danger (ERC >= 90th percentile, area-wide POD of 0.42) exhibited superior skill than RFW forecasts coincident with lesser fire danger (ERC < 90th percentile, area-wide POD of 0.23) (Figure 3.3a). Further, area-wide POD_{SS} coincident with high fire danger was nearly double that of RFW coincident with lesser fire danger. Area-wide POD relative improvement over the reference was 142.1% for LFs with high fire danger and 116.2% for those LFs with lesser fire danger.

POD and POD_{SS} were ubiquitously higher for RFWs coincident with high fire danger than lesser fire danger across WFOs (Figure 3.3a). Similar to findings for differences in skill metrics between human- and lightning-caused fires, we show improvements in POD, FAR, and CSI between RFWs issued during lesser fire danger and during high fire danger across WFOs (Figure 3.3b). For example, we find a POD of 0.61 at for the Pendleton WFO during high fire danger, which is well higher than the POD during lesser fire danger (0.34) and showed a 195% improvement over the reference forecast. We found statistically significant skill for all RFWs conditioned by fire danger except for RFWs issued by PIH coincident with ERC < 90th percentile.

4. Discussion

4.1 Performance in the context of rare event forecasting

We found that RFWs have skill as forecasts for the occurrence of new LFs across the Northwestern US and further demonstrate substantial improvement from reference forecasts (Figure 4.1). This is an important finding that indicates the added value RFWs provide to fire managers and the public. As was expected, overall performance metrics were low due to the rare nature of LFs, and the fact that we constrained our definition of fire activity to new LFs rather than also accounting for growth on existing fires or the number of new ignitions. In addition, forecasts of high-risk, rare events, are prone to hedging, where the cost of a missed forecast exceeds the forecaster's risk tolerance leading to the issuance of more forecasts and a greater number of false alarms (Murphy and Winkler, 1971; Murphy, 1991). Other forecasts of high-risk, rare events such as flash floods (Clark *et al.*, 2014) and earthquakes (Holliday *et al.*, 2005; Shcherbakov *et al.,* 2010) similarly demonstrate the consequences of hedging and lower metrics. By reporting the relative improvement of RFWs alongside skill scores, we mitigate biases introduced through hedging and are able to show that forecasts are considerably more skillful than the reference forecast.

4.2 Lightning- and human-caused fires

A majority (55.6%) of LFs across the study area were human-caused. However, the performance of RFWs for the occurrence of human-caused LFs was quite low with only one WFO having POD > 0.25. Conversely, performance was notably higher for lightning-ignited fires with all WFOs having POD > 0.25 and three WFOs having POD > 0.5. The interaction between lightning and fire occurrence is well understood (Abatzoglou *et al.*, 2016), although there is some debate on which factors are most critical for determining fire ignition potential (Nauslar, 2014). Previous research has shown that the presence of dry thunderstorms, low fuel moistures, and fuel type impact the ignition efficiency of lightning (Rorig and Ferguson, 1999). The probability of dry thunderstorms as agents of fire ignition is often included in the issuance of RFWs. Improvements in dry thunderstorm forecasting and continued research on lightning characteristics (e.g., polarity, residence time, and amplitude) for ignition potential (see Shultz *et al.* (2019)) are likely to increase RFW performance for lightning-caused fires.

By contrast, RFWs do not explicitly include predictors of human ignitions. The lesser skill of RFW for capturing human-caused LFs is consistent with prior research that shows that human-caused fires are more difficult to predict (Martínez *et al.*, 2008; Magnussen and Taylor, 2012) and occur over a more diverse set of fuel moistures and broader period of the year than lightning-caused fires (Balch *et al.*, 2017). Other factors beyond fuels, weather, and topography tend to influence human-caused ignitions such as road, population, and railroad density, day of the week, holidays, and socioeconomic status (Costafreda-Aumedes *et al.*, 2017).

The issuance of RFWs may alter human activity resulting in degraded skill. RFWs for non-lightning events (e.g., hot and dry conditions, high winds) may act as a preventative measure for fire managers and the public that reduce the number of new large human-caused fires. For example, the issuance of RFWs can promote action by local land agencies to restrict campfire usage, limit silvicultural and agricultural burning, and bolster suppression capability in the affected areas. Contrarily, there are claims that illegal burners and arsonists may view RFWs as a window of opportunity to maximize their efforts, although research suggests this may not be well-founded (Mees, 1991). Collectively, these factors highlight reasons for reduced performance and skill of RFWs for human-caused fires. More research is needed here to discern the efficacy of RFWs in limiting human ignitions.

4.3. Fuel dryness as a prerequisite for RFW issuance

We found improved RFW skill for new LFs conditioned on fuel dryness. LFs generally occur when fuels are more receptive due to weather and climate drivers (Barbero *et al.*, 2014; Abatzoglou *et al.*, 2018) and RFWs are intended to consider some measure of fuel dryness. Our results reinforce the added value of RFW forecasts that explicitly integrate objective measures of fuel dryness. For this study, we chose ERC because of its representativeness of seasonal drying trends and widespread usage by regional fire management, although we acknowledge that other fire danger indices may be more appropriate for different geographic areas or times of the year depending on the dominant drivers of fire activity for the expected event. For example, in western Washington, downslope wind events that bring hot and dry winds from the Columbia Basin across the Cascades can occur throughout the year and are a known critical fire weather pattern when they co-occur with dry fuels (Brewer *et al.*, 2012; Werth *et al.*, 2016). These events typically occur on short timescales and thus the fuels response and fire risk may be better resolved by 10-/100-hour fuel moisture values than ERC. Additional studies that examine a variety of fuel aridity metrics throughout the year may aid in the efficacy of RFWs.

4.4 Limitations of assessing and interpreting skill of RFWs

RFWs are issued for areas of relatively homogenous climate and fuels. We caution that comparisons of performance between WFOs need to consider the climate, fuels, and frequency of events that differ markedly across FWZs. Further, WFOs with larger FWZ sizes could show artificially better skill because of a larger pool of fire occurrence. Previous studies have shown similar results where performance increases with the scale of the geographic area considered but tends to result in decreased value to the end-user (Scherbakov *et al.* 2010). Additional analyses that examine the specific criteria for RFW by zones as well as the host of biophysical and human factors may help shed light on differences in forecast skill.

Lastly, we considered RFWs as forecasts of new LF occurrence, although in reality, RFWs may be issued for weather conditions (e.g., atmospheric instability, high winds, low relative humidity) that promote growth on existing fires and heighten the potential for rapid rates of fire spread for new ignitions. Our explicit treatment of RFWs as forecasts for new LFs results in a low estimate of skill as we classify RFW days that may have rapid growth on existing fires but no new LFs as false alarms. Recent geospatial datasets of daily fire incident status reports (SIT-209s, St. Denis *et al.*, 2020) and burned area information from satellite imagery (Andela *et al.*, 2019) aim to provide fire growth information for larger fires and thus may be useful for future evaluation of RFWs and other fire hazard warning systems.

5. Conclusions

We found favorable skill of RFWs for a meaningful measure of fire activity broadly across the Northwestern U.S.. We additionally demonstrated that RFW skill was significantly better for lightning-caused fires and when issued coincident to high fuel dryness . While this is the first known study on RFW skill married to actual fire activity, our measures of skill are specific only to the occurrence of new large fires. Still, these forecasts may have value for fire early warning systems beyond that reported here.

Our results provide a means for discussing the quality of RFWs and highlight recommendations on improving RFWs while preserving value to end-users. The definition of RFWs and criteria used to issue these forecasts should be explicit, centrally-documented, and easily verifiable. We discovered many different WFO interpretations of the RFW definition and found numerous qualitative RFW criteria that would be especially challenging to verify directly. RFW criteria should include a measure of fuel dryness (e.g., ERC, 1000-hr fuel moisture) and be flexible enough to accommodate different weather regimes that drive fire activity throughout the year. Further, improved empirical analyses on the weather and fuels conditions that lead to significant ignitions and rapid rates of spread need to be performed to determine appropriate local RFW criteria, and these criteria should be quantifiable and concise so that performance may be easily assessed and improved upon. Such analyses may draw upon studies that have identified meteorological and fuel moisture thresholds important for ignitions and rapid spread rates (Abatzoglou *et al.*, 2018; Rorig and Ferguson, 1999; Podur and Wotton, 2011).

RFWs and other early warning systems for fire may benefit from incorporating a probabilistic framework that is better suited for high-risk rare event forecasting (Murphy, 1991; Gneiting and Katzfuss, 2014) and quantitative risk assessments (Casati *et al.*, 2008). A probabilistic forecast would be especially useful for the fire community, where decisions are commonly made with high economic cost and human risk factors (Worsnop *et al.*, 2020; Noonan-Wright *et al.*, 2011). The NOAA Forecasting a Continuum of Environment Threats (FACETs) framework seeks to supplement or replace existing NWS deterministic products with high-resolution probabilistic information (Rothfusz *et al.*, 2018) and may be applicable for moving RFWs in this direction. Beyond RFWs, other fire early warning systems should consider providing probabilistic information that identify the range of potential scenarios that would ultimately lead to better, more actionable decisions by end-users.

References

1. Bowman, D.M.; Balch, J.K.; Artaxo, P.; Bond, W.J.; Carlson, J.M.; Cochrane, M.A.; D'Antonio, C.M.; DeFries, R.S.; Doyle, J.C.; Harrison, S.P.; Johnston, F.H.; Keeley, J.E.; Krawchuk, M.A.; Kull, C.A.; Marston, J.; Moritz, M.A.; Prentice, I.C.; Roos, C.I.; Scott, A.C.; Swetnam, T.W.; Werf, G.R.; Pyne, S.J. Fire in the Earth system. *Science*, **2009**, 324, 481–484.
2. Moritz, M.A.; Batllori, E.; Bradstock, R.A.; Gill, A.M.; Handmer, J.; Hessburg, P.F.; Leonard, J.; McCaffrey, S.; Odion, D.C.; Schoennagel, T.; Syphard, A.D., Learning to coexist with wildfire. *Nature*, **2014**, 515, 58–66.
3. Jolly, W.M.; Cochrane, M.A.; Freeborn, P.H.; Holden, Z.A.; Brown, T.J.; Williamson, G.J.; Bowman, D.M. Climate-induced variations in global wildfire danger from 1979 to 2013. *Nat. Commun.*, **2015**, 6, 7537.
4. Balch, J.K.; Bradley, B.A.; Abatzoglou, J.T.; Nagy, R.C.; Fusco, E.J.; Mahood, A.L. Human expansion of the fire niche across the United States. *Proc. Natl. Acad. Sci. USA*, **2017**, 114, 2946–2951.
5. Brewer, M.J.; Clements, C.B. The 2018 Camp Fire: meteorological analysis using in situ observations and numerical simulations. *Atmosphere*, **2020**, 11, 47.
6. Nauslar, N.J.; Abatzoglou, J.T.; Marsh, P.T. The 2017 North Bay and Southern California Fires: A case study. *Fire*, **2018**, 1, 18.
7. Balch, J.K.; Schoennagel, T.; Williams, A.P.; Abatzoglou, J.T.; Cattau, M.E.; Mietkiewicz, N.P.; St. Denis, L.A. Switching on the Big Burn of 2017. *Fire*, **2018**, 1, 17.
8. Barbero, R.; Abatzoglou, J.T.; Larkin, N.K.; Kolden, C.A.; Stocks, B. Climate change presents increased potential for very large fires in the contiguous United States. *Int. J. Wildland Fire*, **2015**, 24, 892–899.
9. Radeloff, V.C.; Helmers, D.P.; Kramer, H.A.; Mockrin, M.H.; Alexandre, P.M.; Bar-Massada, A.; Butsic, V.; Hawbaker, T.J.; Martinuzzi, S.; Syphard, A.D.; Stewart, S.I. Rapid growth of the US wildland-urban interface raises wildfire risk. *Proc. Natl. Acad. Sci. USA*, **2018**, 115, 3314–3319.
10. de Groot, W.J.; Wotton, B.M.; Flannigan, M.D. Wildland fire danger rating and early warning systems. In *Wildfire Risks, Hazards, and Disasters*; Paton, D., Buergelt, P.T.,

McCaffrey, S., Tedim, F., Eds.; Elsevier, Inc.: Amsterdam, Netherlands, 2014; pp. 268, ISBN 978-012-410-434-1.

11. Deeming, J.E.; Burgan, R.E.; Cohen, J.D. The 1978 National Fire-Danger Rating System: Technical documentation. U.S. Department of Agriculture Forest Service Intermountain Forest and Range Experiment Station General Tech. Rep. INT-169, **1983**, pp. 44.
12. Wagner, C.E. Development and structure of the Canadian forest fire index system. Canadian Forestry Service, Headquarters, Ottawa Forestry Tech. Rep. 35, **1987**, pp 35.
13. Jolly, W.M.; Freeborn, P.H.; Page, W.G.; Butler, B.W. Severe Fire Danger Index: A forecastable metric to inform firefighter and community wildfire risk management. *Fire*, **2019**, 2, 47.
14. Rolinski, T.; Capps, S.B.; Fovell, R.G.; Cao, Y.; D'Agostino, B.J.; Vanderburg, S. The Santa Ana Wildfire Threat Index: Methodology and operational implementation. *Wea. Forecasting*, **2016**, 31, 1881–1897.
15. Srock, A.F., Charney, J.J., Potter, B.E., Goodrick, S.L. The Hot-Dry-Windy Index: A new fire weather index. *Atmosphere*, **2019**, 9, 279.
16. Erickson, M.J.; Charney, J.J.; Colle, B.A. Development of a fire weather index using meteorological observations within the Northeast United States. *J. Appl. Meteor. Climatol.*, **2016**, 55, 389–402.
17. NWS, National Weather Service. National Weather Service Instruction 10-401, Fire Weather Services Product Specification, 29 December 2017. pp. 24. Available online: https://www.nws.noaa.gov/directives/sym/pd01004001curr.pdf (accessed on 27 March 2020).
18. Brooks, H.E.; Correia, J. Long-term performance metrics for National Weather Service tornado warnings. *Wea. Forecasting*, **2018**, 33, 1501–1511.
19. Obermeier, H. M.; Anderson, M.R. Verification and analysis of impact-based tornado warnings in the Central Region of the National Weather Service. *Electronic J. Severe Storms Meteor.*, **2014**, 10, 1–20.
20. Clark, R.A.; Gourley, J.J., Flamig, Z.L., Hong, Y.; Clark, E. CONUS-wide evaluation of National Weather Service flash flood guidance products. *Wea. Forecasting*, **2014**, 29, 377–392.

21. McGovern, A.; Gagne, D. J., 2nd; Williams, J. K.; Brown, R. A.; Basara, J. B. Enhancing understanding and improving prediction of severe weather through spatiotemporal relational learning. *Mach. Learn.*, **2014**, 95, 27–50.

22. Werth, P.A.; Potter, B.E.; Alexander, M.E.; Clements, C.B.; Cruz, M.G.; Finney, M.A.; Forthofer, J.M.; Goodrick, S.L.; Hoffman, C.; Jolly, W.M.; McAllister, S.S.; Ottmar, R.D.; Parsons, R.A. Synthesis of knowledge of extreme fire behavior: Volume 2 for fire behavior specialists, researchers, and meteorologists. U.S. Department of Agriculture, Forest Service Pacific Northwest Research Station General Tech. Rep. PNW-GTR-891, **2016**, pp. 258.

23. Short, K.C. Spatial Wildfire Occurrence Data for the United States, 1992–2015 [FPA_FOD_20170508], 4th ed.; Forest Service Research Data Archive, 2017. Available online: https://www.fs.usda.gov/rds/archive/catalog/RDS-2013-0009.4 (accessed on 7 August 2019).

24. Agee, J.K. Fire Ecology of Pacific Northwest Forests, 2nd ed.; Island Press: Washington, D.C., USA, 1996; pp. 505, ISBN 978-155-963-230-0.

25. Abatzoglou, J.T., Kolden, C.A., Balch, J.K., Bradley, B.A. Controls on interannual variability in lightning-caused fire activity in the western US., *Environ. Res. Lett.*, **2016**, 11, 4.

26. IEM, Iowa Environmental Mesonet. Archived NWS watch/warnings. Available online: https://mesonet.agron.iastate.edu/request/gis/watchwarn.phtml (accessed on 30 May 2019).

27. NWS, National Weather Service. National Weather Service Instruction 10-407, Fire Weather Services Zone Change Process, 5 January 2018. pp. 6. Available online: https://www.nws.noaa.gov/directives/sym/pd01004007curr.pdf (accessed on 27 March 2020).

28. Nagy, R.C.; Fusco, E.; Bradley, B.A.; Abatzoglou, J.T.; Balch, J.K. Human-related ignitions increase the number of large wildfires across U.S. ecoregions. *Fire*, **2018**, 1, 4.

29. Broxton, P.D.; Zeng, X.; Sulla-Menashe, D.; Troch, P.A. A global land cover climatology using MODIS data. *J. Appl. Meteor. Climatol.*, **2014**, 53, 1593–1605.

30. Abatzoglou, J.T. Development of gridded surface meteorological data for ecological applications and modelling. *Int. J. Climatol.*, **2013**, 33, 121–131.
31. Freeborn, P.H.; Cochrane, M.A.; Jolly, W.M. Relationships between fire danger and the daily number and daily growth of active incidents burning in the northern Rocky Mountains, USA. *Int. J. Wildland Fire*, **2015**, 24, 900–910.
32. Abatzoglou, J.T.; Kolden, C.A. Relationships between climate and macroscale area burned in the western United States. *Int. J. Wildland Fire*, **2013**, 22, 1003-1020.
33. Barbero, R.; Abatzoglou, J.T.; Steel, E.A. Modeling very large-fire occurrences over the continental United States from weather and climate forcing. *Environ. Res. Lett.,* **2014**, 9, 12.
34. Heinsch, F.A.; Andrews, P.L.; Kurth, L.L. Implications of using percentiles to define fire danger levels. Extended Abstract P1.5. In: Eighth Symposium on Fire and Forest Meteorology, 12-15 October 2009, Kalispell, MT.
35. Dalton, M.M.; Abatzoglou, J.T.; Evers, L.; Hegewisch, K. Projected changes in the energy release component under climate change in northwest predictive services areas, A report to the Bureau of Land Management Oregon-Washington State Office, The Oregon Climate Change Research Institute (OCCRI), 2015, Corvallis, Oregon, USA. pp. 88.
36. Murphy, A.H.; Winkler, R.L. A general framework for forecast verification. *Mon. Weather. Rev.*, **1987**, 115, 1330–1338.
37. Doswell, C.A.; Davies-Jones, R.; Keller, D.L. On summary measures of skill in rare event forecasting based on contingency tables. *Wea. Forecasting*, **1990**, 5, 576–585.
38. Wilks, D.S. Statistical Methods in the Atmospheric Sciences, 3rd ed.; Academic Press: Cambridge, MA, USA; pp. 676, ISBN 978-012-385-022-5.
39. Abatzoglou, J. T.; Balch, J. K.; Bradley, B. A.; Kolden, C. A. Human-related ignitions concurrent with high winds promote large wildfires across the USA. *Int. J. Wildland Fire*, **2018**, 27, 377–386.
40. Vaughan, M.T.; Tang, B.H.; Bosart, L.F. Climatology and analysis of high-impact, low predictive skill severe weather events in the northeast United States. *Wea. Forecasting*, **2017**, 32, 1903–1919.

41. Mason, I.B. Binary events. In Forecast Verification: A Practitioner's Guide in Atmospheric Science, 1st ed.; Jolliffe, I.T., Stephenson, D.B, Eds.; John Wiley & Sons Ltd.: West Sussex, England, UK, 2003; pp. 37–76, ISBN 978-047-086-441-8.
42. Schaefer, J.T. The critical success index as an indicator of warning skill. *Wea. Forecasting*, **1990**, 5, 570–575.
43. Potts, J.M. Basic concepts. In Forecast Verification: A Practitioner's Guide in Atmospheric Science, 1st ed.; Jolliffe, I.T., Stephenson, D.B, Eds.; John Wiley & Sons Ltd.: West Sussex, England, UK, 2003; pp. 13–36, ISBN 978-047-086-441-8.
44. Mason, S.J. On using "climatology" as a reference strategy in the Brier and ranked probability skill scores. *Mon. Weather Rev*., **2004**, 132, 1891–1895.
45. Warner, T.T. Numerical Weather and Climate Prediction, 1st ed.; Cambridge University Press: New York, NY, USA, 2011; pp. 550, ISBN 978-051-176-324-3.
46. Mittermaier, M.P. The potential impact of using persistence as a reference forecast on perceived forecast skill. *Wea. Forecasting*, **2008**, 23, 1022–1031.
47. Murphy, A.H.; Winkler, R.L. Forecasters and probability forecasts: Some current problems. *Bull. Am. Meteorol. Soc.*, **1971**, 52, 239–247.
48. Murphy, A.H. Probabilities, odds, and forecasts of rare events. *Wea. Forecasting*, **1991**, 6, 302–307.
49. Holliday, J.R.; Nanjo, K.Z.; Tiampo, K.F.; Rundle, J.B.; Turcotte, D.L. Earthquake forecasting and its verification. *Nonlinear Process. Geophys.*, European Geosciences Union (EGU), **2005**, 12, 965–977.
50. Shcherbakov, R.; Turcotte, D.L.; Rundle, J.B.; Tiampo, K.F.; Holliday, J.R. Forecasting the locations of future large earthquakes: An analysis and verification. *Pure Appl. Geophys.*, **2010**, 167, 743–749.
51. Nauslar, N.J. Examining the lightning polarity of lightning caused wildfires. In Proceedings of the 23rd International Lightning Detection Conference, 5th International Lightning Detection Conference, Tucson, Arizona USA, 18–21 March 2014; pp.15.
52. Rorig, M.L.; Ferguson, S.A. Characteristics of lightning and wildland fire ignition in the Pacific Northwest. *J. Appl. Meteorol.*, **1999**, 38, 1565–1575.

53. Schultz, C.J.; Nauslar, N.J.; Wachter, J.B.; Hain, C.R.; Bell, J.R. Spatial, temporal and electrical characteristics of lightning in reported lightning-initiated wildfire events. *Fire*, **2019**, 2, 18.
54. Martínez, J.; Vega-Garcia, C.; Chuvieco, E. Human-caused wildfire risk rating for prevention planning in Spain. *J. Environ. Manage.*, **2009**, 90, 1241–1252.
55. Magnussen, S.; Taylor, S.W. Prediction of daily lightning-and human-caused fires in British Columbia. *Int. J. Wildland Fire*, **2012**, 21, 342–356.
56. Costafreda-Aumedes, S.; Comas, Carles; Vega-Garcia, C. Human-caused fire occurrence modelling in perspective: A review. *Int. J. Wildland Fire*, **2017**, 26, 983–998.
57. Mees, R. Is arson associated with severe fire weather in Southern California? *Int. J. Wildland Fire*, **1991**, 1, 97–100.
58. Barbero, R.; Abatzoglou, J.T.; Steel, E.A. Modeling very large-fire occurrences over the continental United States from weather and climate forcing. *Environ. Res. Lett.*, **2014**, 9, 124009.
59. Brewer, M.C.; Mass, C.F.; Potter, B.E. The West Coast thermal trough: Climatology and synoptic evolution. *Mon. Wea. Rev.*, **2012**, 140, 3820–3843.
60. St. Denis, L.A.; Mietkiewicz, N.P.; Short, K.C.; Buckland, M.; Balch, J. K. All-hazards dataset mined from the US National Incident Management System 1999–2014. *Sci. Data*, **2020**, 7**,** 64.
61. Andela, N.; Morton, D.C.; Giglio, L.; Paugam, R.; Chen, Y.; Hantson, S.; van der Werf, G.R.; Randerson, J.T. The Global Fire Atlas of individual fire size, duration, speed and direction, *Earth Syst. Sci. Data*, **2019**, 11, 529–552.
62. Podur, J.; Wotton, B.M. Defining fire spread event days for fire-growth modelling. *Int. J. Wildland Fire*, **2011**, 20, 497–507.
63. Gneiting, T.; Katzfuss, M. Probabilistic forecasting. *Annu. Rev. Stat. Its Appl.*, **2014**, 1, 125–151.
64. Casati, B.; Wilson, L.J.; Stephenson, D.B.; Nurmi, P.; Ghelli, A.; Pocernich, M.; Damrath, U.; Ebert, E.E.; Brown, B.G.; Mason, S. Forecast verification: current status and future directions. *Met. Apps*., **2008**, 15, 3–18.

65. Worsnop, R.P.; Scheuerer, M.; Hamill, T.M. Extended-range probabilistic fire-weather forecasting based on ensemble model output statistics and ensemble copula coupling. *Mon. Wea. Rev.*, **2020**, 148, 499–521.

66. Noonan-Wright, E. K.; Opperman, T.S.; Finney, M.A.; Zimmerman, G.T.; Seli, R.C.; Elenz, L.M.; Calkin, D.E.; Fiedler, J.R. Developing the US Wildland Fire Decision Support System. *J. Combust.* **2011,** 168473.

67. Rothfusz, L.P.; Schneider, R.; Novak, D.; Klockow-McClain, K.; Gerard, A.E.; Karstens, C.; Stumpf, G.J.; Smith, T.M. FACETs: A proposed next-generation paradigm for high-impact weather forecasting. *Bull. Amer. Meteor. Soc.*, **2018**, 99, 2025–2043.

```
URGENT - FIRE WEATHER MESSAGE
NATIONAL WEATHER SERVICE SPOKANE WA
440 AM PDT FRI AUG 28 2015

...RED FLAG WARNING FOR VERY WINDY CONDITIONS AND LOW RELATIVE
HUMIDITIES SATURDAY...


IDZ101-WAZ684-686-687-282300-
/O.NEW.KOTX.FW.W.0020.150829T1800Z-150830T0300Z/
NORTHERN AND CENTRAL IDAHO PANHANDLE (ZONE 101)-
EAST WASHINGTON OKANOGAN/METHOW VALLEYS (ZONE 684)-
EAST WASHINGTON NORTHEAST (ZONE 686)-
EAST WASHINGTON OKANOGAN HIGHLANDS (ZONE 687)-
440 AM PDT FRI AUG 28 2015

...RED FLAG WARNING IN EFFECT FROM 11 AM TO 8 PM PDT SATURDAY FOR
WIND AND LOW RELATIVE HUMIDITY FOR NORTHERN WASHINGTON AND NORTH
IDAHO...

THE NATIONAL WEATHER SERVICE IN SPOKANE HAS ISSUED A RED FLAG
WARNING FOR WIND AND LOW RELATIVE HUMIDITY...WHICH IS IN EFFECT
FROM 11 AM TO 8 PM PDT SATURDAY.

* AFFECTED AREA: FIRE WEATHER ZONE 101 NORTHERN AND CENTRAL
  IDAHO PANHANDLE (ZONE 101)...FIRE WEATHER ZONE 684 EAST
  WASHINGTON OKANOGAN/METHOW VALLEYS (ZONE 684)...FIRE WEATHER
  ZONE 686 EAST WASHINGTON NORTHEAST (ZONE 686) AND FIRE WEATHER
  ZONE 687 EAST WASHINGTON OKANOGAN HIGHLANDS (ZONE 687).

* WINDS: SOUTHWEST 25 TO 35 MPH WITH GUSTS UP TO 60 MPH. GUSTS AS
  HIGH AS 70 MPH IN THE MOUNTAINS.

* RELATIVE HUMIDITIES: 20 TO 30 PERCENT IN THE VALLEYS. 30 TO 40
  PERCENT IN THE MOUNTAINS.

* IMPACTS: VERY WINDY CONDITIONS AND LOW RELATIVE HUMIDITIES WILL
  RESULT IN RAPID GROWTH ON EXISTING FIRES OR ANY NEW IGNITIONS.

PRECAUTIONARY/PREPAREDNESS ACTIONS...

A RED FLAG WARNING MEANS THAT CRITICAL FIRE WEATHER CONDITIONS
ARE EITHER OCCURRING NOW...OR WILL SHORTLY. A COMBINATION OF
STRONG WINDS AND LOW RELATIVE HUMIDITY WILL CREATE EXTREME FIRE
GROWTH POTENTIAL.
```

Figure 1.1: An example of a Red Flag Warning issued by the NWS Spokane forecast office for windy and dry conditions on 28 August 2015.

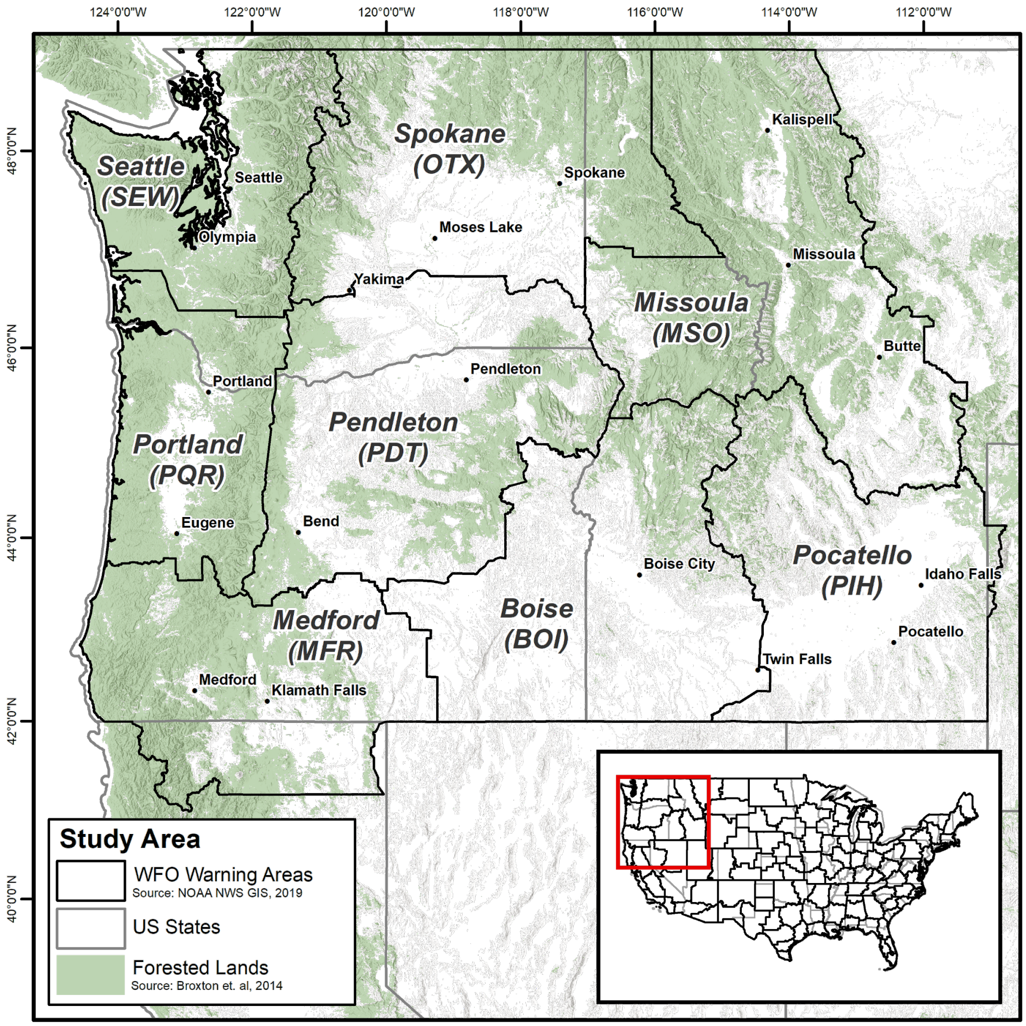

Figure 2.1: Map of the study area with forestlands defined using the 0.5 km MODIS-based global land cover climatology dataset (Broxton *et al.*, 2014) and NWS WFO warning areas based on fire weather zones.

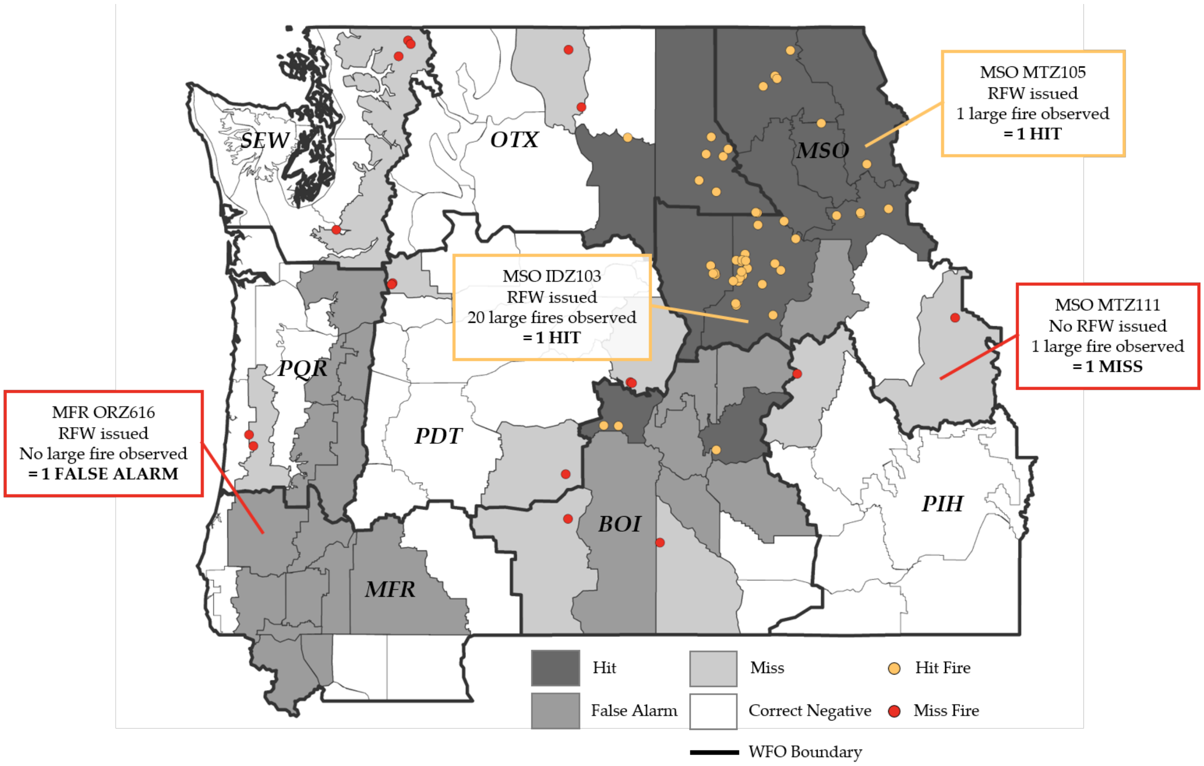

Figure 2.3: An example of the RFW and large fire classification scheme pairing RFWs issued for 10 August 2015 and new large fires from 10–11 August 2015. 32 RFWs were issued resulting in 11 hits, 11 misses, and 21 false alarms across the study area.

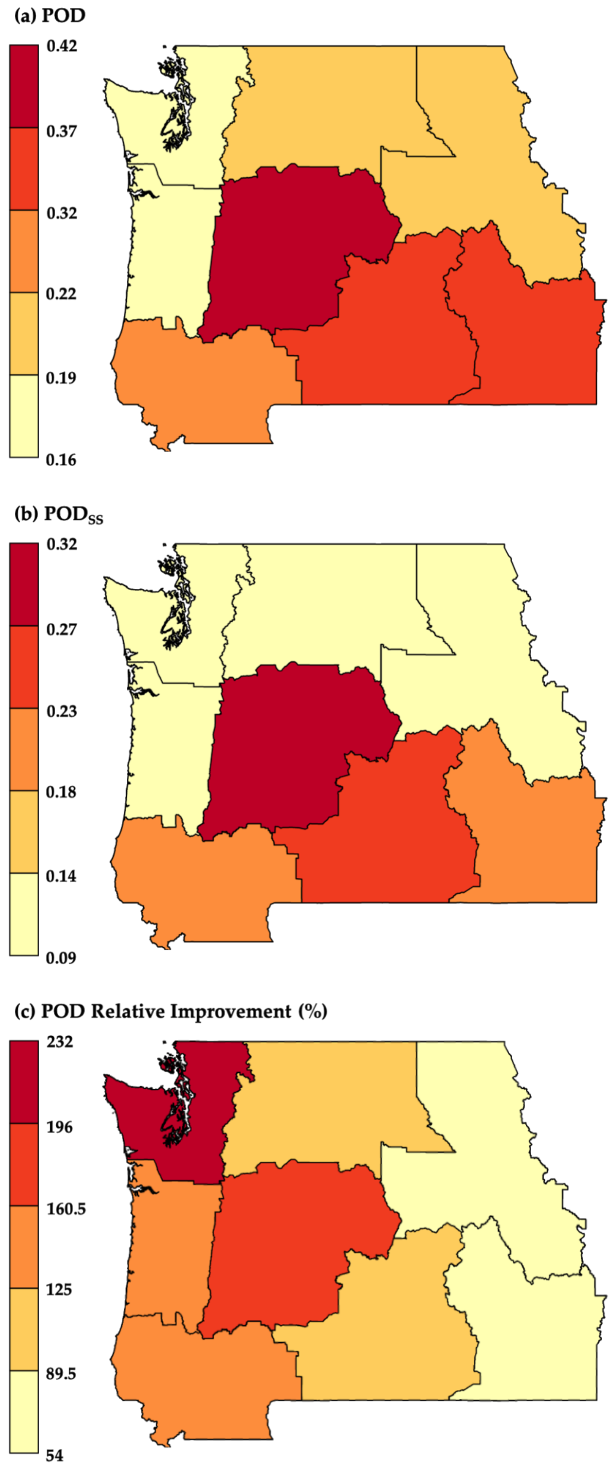

Figure 3.1: Red Flag Warning skill for large fires (>= 90th percentile) for each WFO showing (a) POD, (b) POD_{SS}, and (c) relative improvement over the random climatology reference forecast.

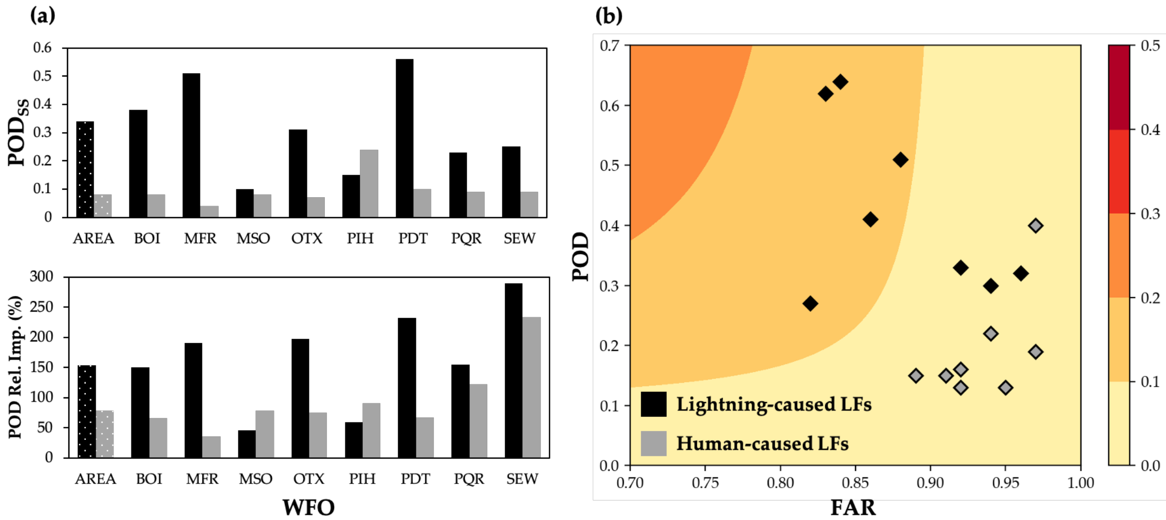

Figure 3.3: POD skill scores and relative improvement (a) and performance metrics (b) of RFWs for lightning- and human-caused large fires across the WFOs. For (b), CSI is found to generally increase with higher POD and decreasing FAR.

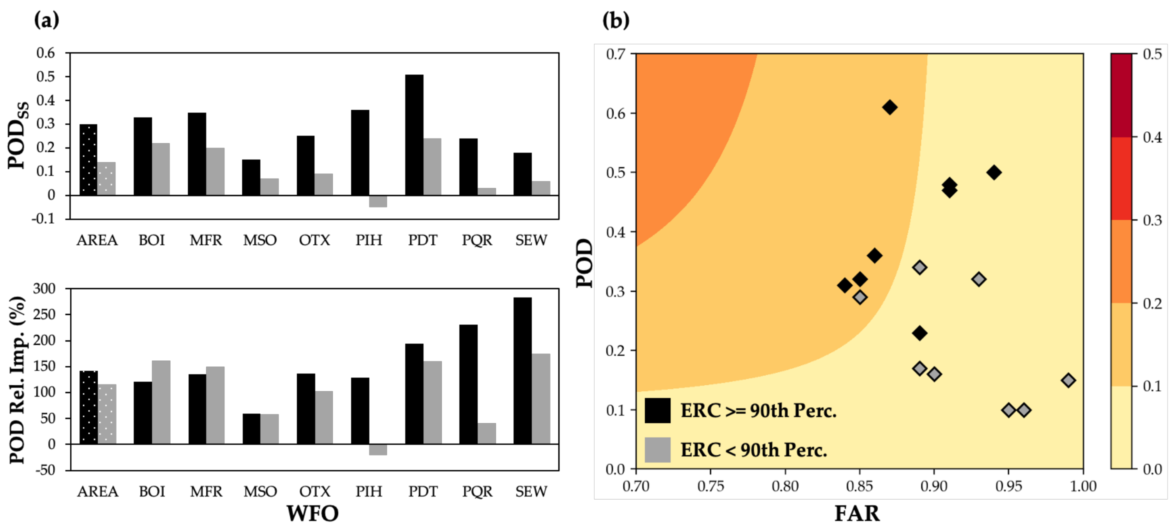

Figure 3.5: POD skill scores and relative improvement (a) and performance metrics (b) of RFWs for when LFs occurred with ERC >= 90th percentile and ERC < 90th percentile values.

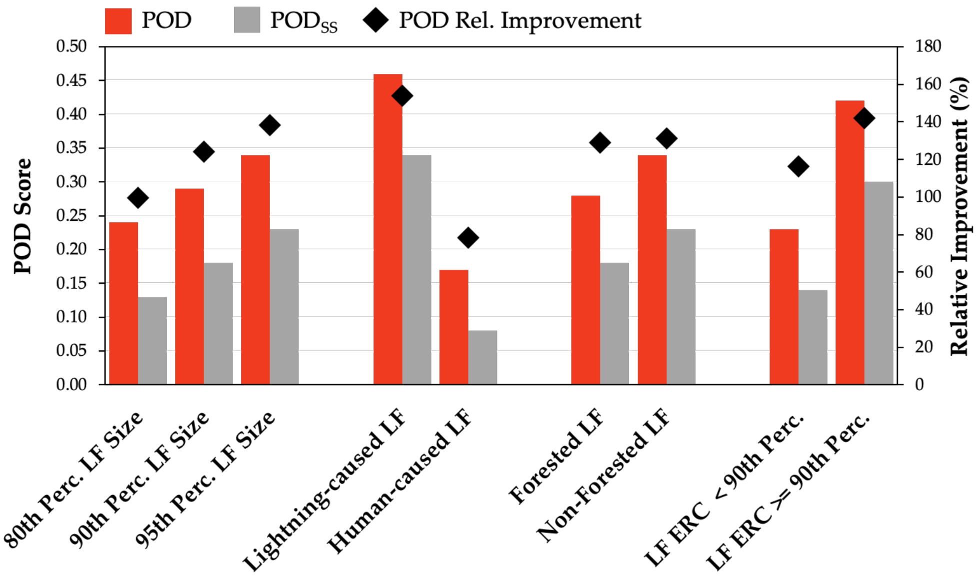

Figure 4.1: Summary of area-wide POD, POD_{SS}, and POD relative improvement for the four classification of LF types.

Table 1.2: Select criteria for issuing RFWs taken from the 2019 Northwest Area Fire Weather Annual Operating Plan.

WFO Pendleton All FWZs	"LIGHTNING: Abundant lightning in conjunction with sufficiently dry fuels (fuels remain dry or critical during and after a lightning event). Warnings are not typically issued for isolated coverage events. Warnings not typically issued for events that will be accompanied by significant rain (greater than 0.25 inches). However, if a lightning event will occur with significant rain, but is then followed by very hot and dry conditions, a warning may be issued if holdover/sleeper fires are a concern."
WFO Portland Zones 605, 607, and 660	"One station (RAWS) must report 35% humidity or less AND 10-minute wind speed of 10 mph AND/OR gusts to 20 mph or more for four hours in an 8- hour block, AND at least TWO other stations reporting 35% humidity or less AND 10- minute wind of 10 mph AND/OR gusts to 20 mph for at least TWO hours. Key RAWS: Horse Creek, Log Creek, Wanderer's Peak, Kosmos, Canyon Creek, Orr Creek, Elk Rock, and 3-Corner Rock. NOTE: Includes stations from zone 659." (Only valid during nighttime hours)
WFO Spokane All FWZs	"An unusually unstable atmosphere. This would be associated with a strong thermal trough which typically forms along the east slopes of the Washington Cascades."

Table 2.2: Forecast performance metrics calculated in this study.

Performance metric	Measures	Equation	Possible range and perfect score
Probability of detection (POD)	Fraction of observed events that were correctly forecast	$POD = \frac{a}{a+c}$	0 to 1 1
False alarm ratio (FAR)	Fraction of predicted events that did not actually occur	$FAR = \frac{b}{a+b}$	0 to 1 0
Critical success index (CSI)	Fraction of correctly forecasted events without consideration to correct negatives	$CSI = \frac{a}{a+b+c}$	0 to 1 1

Table 3.2: Area-wide POD, skill scores, and relative improvement of Red Flag Warnings tested at 80th, 90th, and 95th large fire size thresholds.

	LF Size >= 80th Perc.	**LF Size >= 90th Perc.**	**LF Size >= 95th Perc.**
POD	0.24	0.29	0.34
POD_{SS}	0.13	0.18	0.23
Rel. Imp.	99.4%	124.1%	138.2%

Table 3.4: As in Table 3.1 but for forested and non-forested LF land cover types.

	Forested LF	Non-Forested LF
POD	0.28	0.34
POD_{SS}	0.18	0.23
Rel. Imp.	129.0%	131.3%

Appendix 1: Supplemental Material

Table A1.1a: Complete RFW performance statistics, skill scores, and relative improvement values for large fires >= 80th percentile sizes.

	Forecast Performance			Reference Performance			Skill Score			Rel. Improvement (%)		
	POD	**FAR**	**CSI**	**POD**	**FAR**	**CSI**	**POD**	**FAR**	**CSI**	**POD**	**FAR**	**CSI**
AREA	0.29	0.82	0.13	0.13	0.92	0.05	0.18	0.11	0.08	124.08	-11.3	142.0
BOI	0.37	0.85	0.12	0.17	0.94	0.05	0.25	0.09	0.07	124.78	-8.8	144.6
MFR	0.32	0.79	0.15	0.13	0.91	0.06	0.21	0.14	0.10	135.93	-13.7	156.8
MSO	0.22	0.76	0.13	0.14	0.84	0.08	0.09	0.10	0.05	53.91	-9.7	58.5
OTX	0.22	0.79	0.12	0.10	0.90	0.05	0.13	0.12	0.07	110.62	-12.4	122.8
PIH	0.35	0.93	0.06	0.20	0.96	0.03	0.19	0.03	0.03	73.84	-3.2	82.9
PDT	0.42	0.78	0.17	0.16	0.92	0.05	0.32	0.16	0.12	171.73	-15.9	214.9
PQR	0.19	0.81	0.10	0.08	0.92	0.04	0.12	0.12	0.07	139.30	-12.2	155.2
SEW	0.16	0.86	0.08	0.05	0.96	0.02	0.12	0.10	0.06	231.45	-10.4	254.6

Table A1.1b: Complete RFW performance statistics, skill scores, and relative improvement values for large fires >= 90th percentile sizes.

	Forecast Performance			Reference Performance			Skill Score			Rel. Improvement (%)		
	POD	**FAR**	**CSI**	**POD**	**FAR**	**CSI**	**POD**	**FAR**	**CSI**	**POD**	**FAR**	**CSI**
AREA	0.24	0.72	0.15	0.12	0.86	0.07	0.13	0.16	0.08	99.4	-16.5	113.5
BOI	0.32	0.78	0.15	0.15	0.89	0.07	0.19	0.13	0.09	108.5	-12.9	124.4
MFR	0.26	0.67	0.17	0.13	0.84	0.08	0.16	0.20	0.10	105.9	-20.1	122.4
MSO	0.18	0.61	0.14	0.12	0.73	0.09	0.06	0.16	0.05	46.3	-15.7	51.0
OTX	0.16	0.69	0.12	0.10	0.81	0.07	0.07	0.15	0.06	68.1	-15.0	75.3
PIH	0.36	0.86	0.11	0.19	0.93	0.05	0.21	0.07	0.06	87.2	-7.3	103.2
PDT	0.34	0.67	0.20	0.14	0.87	0.07	0.23	0.23	0.14	139.6	-22.6	171.4
PQR	0.15	0.70	0.11	0.07	0.86	0.05	0.08	0.18	0.06	104.9	-17.9	117.7
SEW	0.13	0.78	0.09	0.05	0.92	0.03	0.09	0.15	0.06	188.7	-15.4	206.8

Table A1.1c: Complete RFW performance statistics, skill scores, and relative improvement values for large fires >= 95th percentile sizes.

	Forecast Performance			Reference Performance			Skill Score			Rel. Improvement (%)		
	POD	**FAR**	**CSI**	**POD**	**FAR**	**CSI**	**POD**	**FAR**	**CSI**	**POD**	**FAR**	**CSI**
AREA	0.34	0.88	0.09	0.14	0.95	0.04	0.23	0.07	0.06	138.2	-7.2	157.4
BOI	0.40	0.92	0.07	0.17	0.97	0.03	0.28	0.05	0.05	136.6	-5.1	152.3
MFR	0.36	0.87	0.11	0.15	0.95	0.04	0.25	0.08	0.07	138.2	-8.2	158.5
MSO	0.25	0.85	0.10	0.15	0.91	0.06	0.11	0.06	0.04	61.3	-5.9	65.1
OTX	0.29	0.85	0.11	0.12	0.94	0.04	0.20	0.10	0.07	153.0	-10.0	174.0
PIH	0.36	0.96	0.04	0.21	0.98	0.02	0.19	0.02	0.02	74.7	-1.7	78.8
PDT	0.51	0.85	0.13	0.18	0.95	0.04	0.40	0.11	0.09	183.4	-10.5	237.2
PQR	0.23	0.88	0.09	0.08	0.96	0.03	0.16	0.08	0.06	191.8	-8.3	217.3
SEW	0.15	0.93	0.05	0.05	0.98	0.02	0.10	0.04	0.03	189.0	-4.4	203.0

Table A1.2a: Complete RFW performance statistics, skill scores, and relative improvement values for lightning-caused large fires >= 90th percentile sizes.

	Forecast Performance			Reference Performance			Skill Score			Rel. Improvement (%)		
	POD	**FAR**	**CSI**	**POD**	**FAR**	**CSI**	**POD**	**FAR**	**CSI**	**POD**	**FAR**	**CSI**
AREA	0.46	0.87	0.11	0.18	0.95	0.04	0.34	0.09	0.08	154.0	-8.7	182.9
BOI	0.51	0.88	0.11	0.20	0.95	0.04	0.38	0.08	0.07	150.4	-7.9	180.5
MFR	0.62	0.83	0.16	0.21	0.95	0.05	0.51	0.13	0.12	190.8	-12.7	245.4
MSO	0.27	0.82	0.12	0.19	0.88	0.08	0.10	0.06	0.04	45.4	-6.0	48.9
OTX	0.41	0.86	0.11	0.14	0.96	0.03	0.31	0.10	0.08	197.6	-9.8	230.5
PIH	0.32	0.96	0.04	0.20	0.98	0.02	0.15	0.02	0.02	59.2	-1.6	66.0
PDT	0.64	0.84	0.15	0.19	0.96	0.04	0.56	0.13	0.12	232.4	-12.7	311.8
PQR	0.33	0.92	0.07	0.13	0.97	0.03	0.23	0.05	0.05	154.3	-5.3	170.6
SEW	0.30	0.94	0.05	0.08	0.99	0.01	0.25	0.04	0.04	289.6	-4.4	325.0

Table A1.2b: Complete RFW performance statistics, skill scores, and relative improvement values for human-caused large fires >= 90th percentile sizes.

	Forecast Performance			Reference Performance			Skill Score			Rel. Improvement (%)		
	POD	**FAR**	**CSI**	**POD**	**FAR**	**CSI**	**POD**	**FAR**	**CSI**	**POD**	**FAR**	**CSI**
AREA	0.17	0.94	0.04	0.09	0.97	0.02	0.08	0.03	0.02	78.4	-2.5	80.7
BOI	0.19	0.97	0.03	0.11	0.98	0.02	0.08	0.01	0.01	66.4	-1.1	66.7
MFR	0.13	0.95	0.04	0.10	0.96	0.03	0.04	0.01	0.01	36.1	-1.2	33.7
MSO	0.16	0.92	0.05	0.09	0.96	0.03	0.08	0.03	0.03	78.4	-3.4	78.3
OTX	0.15	0.91	0.06	0.08	0.95	0.03	0.07	0.04	0.03	75.1	-3.8	79.4
PIH	0.40	0.97	0.03	0.21	0.99	0.01	0.24	0.01	0.01	91.2	-1.4	101.9
PDT	0.22	0.94	0.05	0.13	0.97	0.03	0.10	0.02	0.02	67.7	-2.4	72.7
PQR	0.15	0.89	0.07	0.07	0.95	0.03	0.09	0.06	0.04	122.1	-6.5	133.7
SEW	0.13	0.92	0.05	0.04	0.97	0.02	0.09	0.06	0.04	233.1	-6.1	248.0

Table A1.3a: Complete RFW performance statistics, skill scores, and relative improvement values for forested large fires >= 90th percentile sizes.

	Forecast Performance			Reference Performance			Skill Score			Rel. Improvement (%)		
	POD	**FAR**	**CSI**	**POD**	**FAR**	**CSI**	**POD**	**FAR**	**CSI**	**POD**	**FAR**	**CSI**
AREA	0.28	0.90	0.08	0.12	0.96	0.03	0.18	0.06	0.05	129.0	-6.0	140.7
BOI	0.44	0.96	0.04	0.23	0.98	0.02	0.26	0.02	0.02	87.4	-1.9	94.4
MFR	0.33	0.85	0.11	0.13	0.94	0.04	0.23	0.10	0.08	162.3	-9.8	183.7
MSO	0.22	0.83	0.11	0.15	0.88	0.07	0.09	0.06	0.04	50.6	-6.2	53.5
OTX	0.23	0.87	0.09	0.10	0.95	0.04	0.15	0.08	0.06	137.9	-7.7	149.8
PIH	0.29	0.99	0.01	0.18	0.99	0.01	0.13	0.00	0.00	57.9	-0.5	66.5
PDT	0.45	0.90	0.09	0.15	0.97	0.03	0.35	0.07	0.06	193.7	-6.9	229.7
PQR	0.19	0.85	0.09	0.08	0.93	0.04	0.11	0.09	0.06	125.8	-9.3	140.9
SEW	0.18	0.86	0.08	0.05	0.96	0.02	0.14	0.10	0.06	269.5	-10.4	298.5

Table A1.3b: Complete RFW performance statistics, skill scores, and relative improvement values for non-forested large fires >= 90th percentile sizes.

	Forecast Performance			Reference Performance			Skill Score			Rel. Improvement (%)		
	POD	**FAR**	**CSI**	**POD**	**FAR**	**CSI**	**POD**	**FAR**	**CSI**	**POD**	**FAR**	**CSI**
AREA	0.34	0.90	0.09	0.15	0.96	0.03	0.23	0.06	0.05	131.3	-6.4	150.6
BOI	0.37	0.88	0.10	0.15	0.95	0.04	0.26	0.07	0.06	148.4	-7.3	167.8
MFR	0.36	0.90	0.08	0.16	0.96	0.03	0.24	0.06	0.05	120.6	-5.6	136.5
MSO	0.25	0.90	0.08	0.14	0.95	0.04	0.13	0.05	0.04	84.9	-4.9	91.5
OTX	0.23	0.89	0.08	0.11	0.95	0.04	0.13	0.06	0.05	103.3	-6.2	114.1
PIH	0.37	0.94	0.05	0.21	0.97	0.03	0.20	0.03	0.02	73.0	-2.6	80.4
PDT	0.46	0.84	0.14	0.17	0.95	0.04	0.35	0.11	0.10	178.3	-11.5	227.2
PQR	0.22	0.95	0.04	0.08	0.99	0.01	0.15	0.03	0.03	185.6	-3.1	209.1
SEW	0.05	0.99	0.00	0.05	0.99	0.00	0.00	0.00	0.00	-1.2	0.0	-0.7

Table A1.4a: Complete RFW performance statistics, skill scores, and relative improvement values for forested large fires >= 90th percentile sizes where ERC >= 90th percentile values.

	Forecast Performance			**Reference Performance**			**Skill Score**			**Rel. Improvement (%)**		
	POD	**FAR**	**CSI**	**POD**	**FAR**	**CSI**	**POD**	**FAR**	**CSI**	**POD**	**FAR**	**CSI**
AREA	0.42	0.89	0.10	0.17	0.96	0.04	0.30	0.07	0.06	142.1	-7.2	169.3
BOI	0.47	0.91	0.08	0.21	0.96	0.03	0.33	0.05	0.05	120.6	-5.1	142.3
MFR	0.48	0.91	0.09	0.20	0.96	0.03	0.35	0.06	0.05	135.1	-5.9	160.4
MSO	0.31	0.84	0.12	0.20	0.90	0.07	0.15	0.07	0.05	60.0	-6.6	65.8
OTX	0.36	0.86	0.11	0.15	0.94	0.04	0.25	0.09	0.07	136.5	-9.0	161.9
PIH	0.50	0.94	0.05	0.22	0.98	0.02	0.36	0.04	0.03	129.0	-3.6	156.8
PDT	0.61	0.87	0.12	0.21	0.96	0.03	0.51	0.10	0.09	194.7	-9.9	264.1
PQR	0.32	0.85	0.11	0.10	0.96	0.03	0.24	0.11	0.09	230.9	-11.4	275.1
SEW	0.23	0.89	0.08	0.06	0.97	0.02	0.18	0.09	0.06	282.8	-8.6	319.7

Table A1.4b: Complete RFW performance statistics, skill scores, and relative improvement values for large fires >= 90th percentile sizes where ERC < 90th percentile values.

	Forecast Performance			**Reference Performance**			**Skill Score**			**Rel. Improvement (%)**		
	POD	**FAR**	**CSI**	**POD**	**FAR**	**CSI**	**POD**	**FAR**	**CSI**	**POD**	**FAR**	**CSI**
AREA	0.23	0.91	0.07	0.11	0.96	0.03	0.14	0.05	0.04	116.2	-5.1	124.7
BOI	0.32	0.93	0.06	0.12	0.97	0.02	0.22	0.05	0.04	161.2	-4.7	176.5
MFR	0.29	0.85	0.11	0.12	0.94	0.04	0.20	0.10	0.07	149.4	-9.8	169.7
MSO	0.17	0.89	0.07	0.11	0.93	0.04	0.07	0.04	0.03	57.9	-4.1	59.5
OTX	0.16	0.90	0.07	0.08	0.95	0.03	0.09	0.05	0.03	102.0	-5.2	106.1
PIH	0.15	0.99	0.01	0.19	0.98	0.02	-0.05	0.00	0.00	-20.1	0.4	-23.5
PDT	0.34	0.89	0.09	0.13	0.96	0.03	0.24	0.07	0.06	160.7	-7.5	185.3
PQR	0.10	0.95	0.04	0.07	0.96	0.03	0.03	0.02	0.01	40.6	-1.6	40.6
SEW	0.10	0.96	0.03	0.04	0.99	0.01	0.06	0.02	0.02	174.1	-2.5	180.0