

1 Short-Range Prediction of Wildfire Intensity with Random Forest Model using Weather Forecast
2 and Satellite Fire Radiative Power Observations

3

4 Authors: Christina Kumler-Bonfanti^{1, 2}, Ravan Ahmadov^{1, 2}, Eric James^{1, 2}, and Jebb Q. Stewart²

5

6 September 2022

7

8 ¹ CIRE, University of Colorado, Boulder, CO 80309, USA

9 ² NOAA Global Systems Laboratory, Boulder, CO 80305, USA

10

11 Corresponding author: Christina Kumler, Christina.E.Kumler@noaa.gov

12

13
14

Abstract:

15 Fire Radiative Power (FRP) detection of wildland fires is important in calculating a variety of
16 variables, such as smoke emissions, which are necessary for fire weather and air quality
17 simulations. FRP is derived from satellite observations of radiative energy emitted by fires. In this
18 work, FRP values collected by geostationary and polar-orbiting satellites in 2018 are used in
19 combination with weather forecast model output to derive a forecasted hourly FRP value unique
20 to that model gridcell. Random Forest (RF) models were trained with inputs including day-before
21 averaged FRP satellite values, hour of day, location, temperature, wind speed, and relative
22 humidity to produce an hourly FRP value on the Rapid Refresh (RAP) model resolution. The
23 CONUS region was used as a training domain and a subdomain of longitudes west of 105W were
24 selected to train separate models to better model hourly wildfire FRP. Additionally, models were
25 trained on separate satellite FRP inputs given the difference in satellite sensor resolutions; one
26 used Geostationary Operational Environmental Satellite (GOES) FRP inputs, and the other used a
27 collection of polar orbiting satellite FRP. Overall, mean absolute errors (MAE) for all RF models
28 were lower than the current persistence approach method used by the High Resolution Rapid
29 Refresh (HRRR). In general, MAE was also lower for wildfire case studies not included in model
30 training and testing, with a few extreme exceptions. This method is a good alternative to the
31 existing method and provides a uniquely forecasted hourly FRP based on weather and
32 time/location.

33
34

Significance Statement:

35 This work introduces a type of machine learning model, Random Forest (RF), to the current
36 method in which hourly fire radiative power (FRP) is modeled using a day-before average FRP
37 and hour of day value. By using a RF, numerical weather model information is used as an input
38 with both the day before average FRP as well as the time of day to create a grid and weather
39 specific hourly FRP forecast.

40
41

1. Introduction:

42 Wildfires are an increasing threat around the country and world, and the ability to predict
43 wildfire behavior has become an important area of research. Wildfire modeling and forecasting

44 can be approached from several different angles, such as modeling wildfire spread, fire weather,
45 fuel conditions, ignition, emissions, and smoke etc. [Wooster 2005, Dennidon 2006, Joseph
46 2019, Wiedinmyer 2011, Ager 2021, Linn 2002, Oliveria 2021, Zou 2019, Bakhshaii 2019]
47 Research into different topics involving wildfires continues to increase as well [Wiggins 2020,
48 Dennison 2014, Westerling 2016]. In particular to topics such as fire burning intensity, behavior,
49 and emissions, using satellite derivations of Fire Radiative Power (FRP) is an approximate
50 measure of the heat energy generated by a fire and has proved to be a valuable tool in the
51 modeling process [Li 2018, Vermote 2009, Roberts 2005]. However, FRP detection and
52 modeling are not straightforward. Sometimes, satellite retrieval of these variables is blocked due
53 to clouds or the fire's own dense smoke plume. Satellite retrieval algorithms can also mistakenly
54 register flares (from things such as solar panels at high noon) as FRP.

55 Fire observations from NOAA's operational environmental satellites are available from
56 geostationary and polar orbits. The sensors on these satellites differ in resolution and in the case
57 of polar orbiters, which will only pass over some same spot a few times a day, with the number
58 of observations of a particular location increasing with latitude. FRP is a proxy for fire intensity
59 measurement, and the typical range of values will differ depending on whether it was retrieved
60 by one type of satellite or the other. For instance, GOES takes more frequent FRP measurements
61 but at lower resolutions and represents a larger area [Xu 2010]. This larger area may produce a
62 larger magnitude of FRP due to the larger amount of active burning within the pixel footprint.
63 Conversely, polar orbiters take FRP measurements at much higher resolutions and therefore each
64 FRP measurement is usually smaller in value. Because of these differences, it was better to
65 model FRP to a specific grid resolution for different satellite types.

66 Another challenge is predicting the fire intensity in real time, which is further
67 complicated with both the complexity of the fire behavior processes and human intervention (ex:
68 firefighting operations). Given the lag in satellite FRP retrievals as well as potential for false
69 detections, it cannot be assumed that previous-hour FRP will be available when modeling future
70 hourly FRP. Unfortunately, this is not ideal in all situations, such as destructive wildfires that last
71 shorter than a day. However, in modeling smoke and aerosols most fires that last longer than a
72 day can be represented with a more consistent source of daily mean FRP from the satellites.
73 Because of some of these complications, certain data restrictions are applied to the input FRP
74 data in the designed models to target wildfire modeling specifically. However, these

75 complications also indicate how important modeling hourly FRP is for smoke, aerosols, and
76 more.

77

78 *A) Methods using machine learning in fire weather-*

79 Wildfires have been increasing in quantity and destruction costs over the years [Turco
80 2018]. There is an upward trend in the number of acres burned per year due to wildfires. This, in
81 combination with the recent developments and interest in machine learning (ML), has led to an
82 increase in research using ML methods in fire modeling, management, and forecasting. One
83 study found that there has been a steady increase in research as far back as the 1990s and
84 summarizes ML applications in six different aspects of fire management and modeling [Jain
85 2020].

86 ML continues to be a valuable tool in an increasing number of applications. While often
87 considered a black box, ML methods can produce well-performing models without much
88 description about what happened “under the hood.” There is an increase in methods to demystify
89 some of the less obvious parts of machine learning in particular to using neural networks [Scott
90 2021, Samudrala 2018, McGovern 2019] as well as make advancements on explainable and
91 trustable AI [Mamalakis 2022,]. These deep learning methods for machine learning require
92 greater numbers of diverse data to train good models as well as having a high magnitude of
93 parameters both tunable and under-the-hood [LeCun 2015, Gu 2016]. A Random Forest (RF)
94 model can be considered a simpler type of machine learning given the more basic decision-
95 making tree structure as well as more easily interpreted [Breiman 2001]. While not exclusive,
96 three factors led to the selection of a RF model for this project. 1) The limited number of unique
97 FRP measurements per model pixel, meaning the limited number of individual and unique fires
98 in different times and locations, that met conditionals for model training. Additionally, previous
99 hourly FRP was not an input but rather a mean of the day-before satellite FRP. This restriction is
100 due to the current input source and pipeline in the numerical weather prediction (NWP) process.
101 2) The model structure’s unique variable input interpretability and simplicity in structure. 3)
102 Initial experiments with neural networks and other structures did not outperform the RF
103 experiments, but were comparable or performed worse than RF. This research sought to design
104 and test a RF with a small selection of meteorological and FRP inputs to model next-hour FRP,
105 improving upon a persistence-only method used in NWP models.

106

107 *B) HRRR-Smoke's method of calculating FRP-*

108 In the previous section, difficulties in measuring and using FRP in real-time were listed.
109 However, FRP is critical to understanding active fire behavior and subsequent applications such
110 as aerosol transport in the High-Resolution Rapid Refresh (HRRR) coupled with Smoke (HRRR-
111 Smoke) model [Ahmadov 2017]. The current method of calculating hourly FRP uses a simple
112 diurnal cycle based on the 24 hour averaged day-before FRP from polar orbiting satellite data
113 (MODIS and VIIRS) based on the location of the wildfire. The mean FRP is multiplied by
114 whatever the value is on the Gaussian curve based on the hour of day. The reason for this
115 climatological curve shape is because FRP tends to be higher in the afternoons and evenings and
116 lower in the early morning and at night. This is restrictive because it is a persistence method that
117 does not use current weather information or specific location. The behavior of this existing
118 method is predictable and Gaussian by design, but analysis of individual hours shows a wide
119 spread, or range, of hourly FRP forecasts both over and under in predicted value.

120 The goal of using ML to model hourly FRP is to use the same input of yesterday's mean
121 FRP as the existing method, but to also introduce additional information from weather models as
122 well as location to model more unique hourly FRP to fire grid points. The hypothesis is that this
123 will be an improvement upon the current method, which does not account for weather in the
124 hourly FRP equation, and will provide hourly forecasts that deviate from the strict Gaussian
125 structure of the existing method.

126

127 *C) Research Problem and Layout of Paper-*

128 For improving the current persistence method that the HRRR-Smoke model uses to
129 estimate hourly FRP, RF machine learning models were trained using meteorological
130 information as inputs in addition to the previous day's mean FRP. Given the current method of
131 using yesterday's mean FRP, a restriction placed on which FRP inputs could be used to train the
132 RF models. The forecast models need to make assumptions in predicting the fire emissions.
133 These models were trained and tested using a satellite FRP database for both polar orbiting and
134 geostationary satellites from 2018. While the inputs were restricted, the truth labels were the
135 observed hourly FRP values from that database.

136 The remaining sections in the paper will be presented in the following way. Section 2 will
137 describe the weather and satellite inputs used by the machine learning model as well describe
138 how and why these inputs were gridded to the RAP domain. Section 3 describes the RF structure
139 in detail and why that type of ML was selected for the hourly FRP models. Additionally, the
140 section will contain a general description on variable input importance and how those are easily
141 obtained with the RF. The results, section 4, will be divided into two parts where the first
142 subsection will describe the models' performance on the testing dataset and the second will show
143 results from famous wildfires. Lastly, section 5 will summarize the work and models that were
144 trained to produce an hourly FRP value and list the next steps for the RF models' development
145 and use.

146

147 **2. Data Inputs:**

148 *A) Weather Model Inputs-*

149 The Rapid Refresh (RAP) numerical weather prediction model, first implemented
150 operationally in 2012, is a 13-km grid spacing hourly cycling prediction system for North
151 America. The successor of the Rapid Update Cycle (RUC) [Benjamin 2004], the RAP takes
152 advantage of frequent three-dimensional atmosphere observations from commercial aircraft
153 (James et al. 2020) to create an hourly analysis of the atmosphere which is then used for weather
154 prediction up to 21-54 hours. The RAP also features a stratiform cloud/hydrometeor analysis
155 [Benjamin 2021], which uses cloud observations from ceilometer and cloud-top pressure and
156 temperature retrieved from geostationary satellites to create a realistic three-dimensional cloud
157 field in the analysis. The RAP also uses a “moderately-coupled” land-soil data assimilation
158 technique to extend the influence of lower atmospheric observations into the soil state through
159 low level temperature and moisture analysis increments [Benjamin 2022]. Additional details on
160 the RAP and its configuration, as well as an evaluation of its performance for surface weather
161 prediction, is provided by Benjamin et al. [2016].

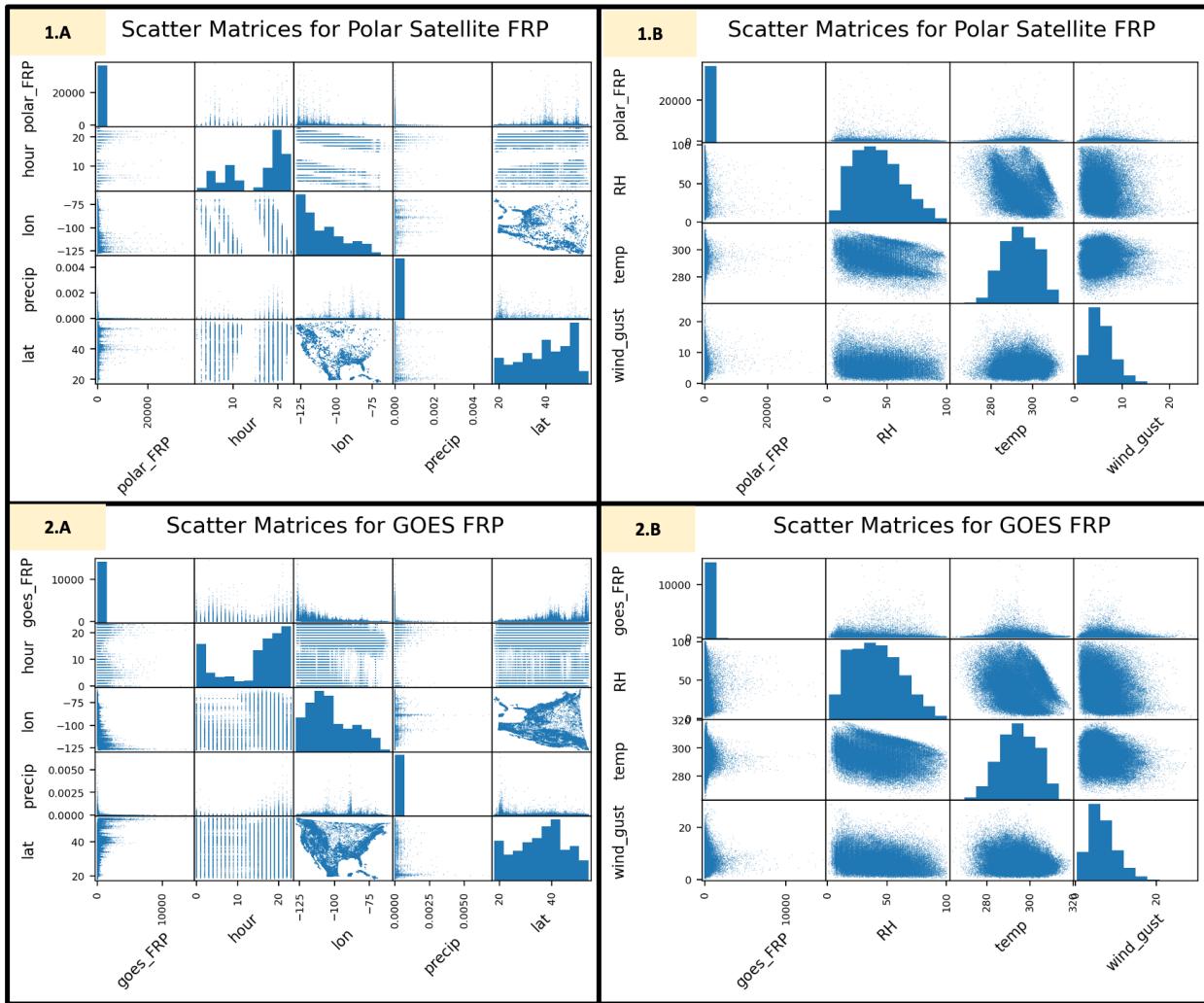
162 FRP from satellites were gridded to the CONUS domain of the RAP model and the
163 corresponding RAP weather inputs paired with the non-zero FRP grid points. Currently, HRRR-
164 Smoke model uses the same Gaussian persistence model to model FRP for all points in the
165 weather model grid based on time zones and time of day. However, FRP is unique and will
166 behave differently based on current and future weather conditions. Because of this, each hourly

167 FRP gridpoint value will behave differently based on different conditions and it was expected
168 that with the addition of weather information, machine learning modeled hourly FRP would
169 perform better than the existing modeling method. By having a consistent FRP and weather
170 domain, the RF models could be trained on daily average satellite FRPs and corresponding
171 hourly meteorological inputs unique to that fire location on the RAP grid.

172 RAP weather model output was collected from 2018 and merged with yesterday mean
173 FRP, which is the mean FRP value over the day before for the same location, for each RAP grid
174 cell where there was a non-zero hourly FRP value present. Summer 2018 was an intensive fire
175 season and provided many hourly FRP values from a fairly large quantity of unique wildfires.
176 RAP current hourly weather variables were used to represent that corresponding hour's
177 meteorological variables. There are weather factors known to impact active wildfire behavior,
178 and a sample subset of weather variables were selected to be used as inputs into hourly FRP
179 modeling to improve upon the persistence method, which does not use weather or location
180 information at all. By keeping the selection reduced, chances of highly correlated variables were
181 limited, which is important when applying explainable ML techniques to models. Additionally,
182 by selecting variables specific to weather models, the models can eventually be integrated into
183 the modeling pipeline and update current hourly FRP calculations. Below is the list of the
184 extracted RAP weather variables selected for experimentation in training the RF models:

- 185 - 2 m temperature (K)
- 186 - Relative humidity (%)
- 187 - 10 m wind gust (m/s)
- 188 - Vegetation type
- 189 - Instantaneous Precipitation (kg m⁻²)

190



191

192 Figure 1: Scatter matrices for all RAP-gridded hourly FRP points from polar satellites over the Conus domain
 193 where the hourly FRP had to be greater than zero. This figure illustrates relationships and potential correlations
 194 between hourly FRP as well as location, time, and RAP weather variables. Panel 1 shows relationships
 195 between variables for the polar orbiting satellite points, and panel 2 shows the same but for GOES. A and B
 196 panel plots are provided to. The variables included are: polar or GOES hourly FRP, respectively, longitude
 197 (lon [degrees]), latitude (lat [degrees]), hour of day hour[UTC]), model precipitation (precip [kg m⁻²]),
 198 temperature (temp [K]), and wind gust (wind_gust [m s⁻¹]).

199

200 The scatter plots in Figure 1 show the relationship between some of the selected input
 201 variables to each other and hourly FRP. These were done for both GOES and polar orbiters on
 202 both the CONUS and western domains, but the plots look similar for both satellite types. One
 203 thing of note in Figure 1, panels 1.A and 1.B, is that the polar orbiter shows gaps in missing FRP
 204 measurements at certain locations during certain times of day in hourly FRP measurements. This

205 introduces a slight bias in how the RF models are trained, since there are only certain times of
206 day that the hourly label is present. These scatter plots also show a slight relationship to the
207 magnitude of FRP and time of day. FRP is known to be higher in the afternoon/evening and
208 lowest at night [Andela 2015]. Correlation is something to factor when selecting input variables
209 into ML models because it impacts the explainability of models' input variables. Ideally,
210 variables are little to not correlated for the most interpretable input importance. To check for
211 high correlation between any of the input variables, Pearson's and Spearman's correlation
212 calculations were used and it was found that the highest Spearman's value was between
213 longitude and polar orbiting FRP, with a value of -0.641 and then -0.53 for the GOES satellites.
214 The longitude relationship might be due to the relationship of wildfires occurring in certain
215 regions and/or due to longer-lasting wildfires in similar locations. This value is not significant
216 enough to be an issue.

217

218 *B) Satellite Data-*

219 Separating polar orbiting and geostationary satellites is important. Polar orbiting satellites
220 collect high resolution data but are not in a fixed orbit, and only pass over an area a few times a
221 day. Conversely, geostationary satellites are stationary and collect data over the same section of
222 the earth at a lower resolution. Because the two satellite groups have different resolutions and
223 sensors, it affects the magnitude of an FRP value and FRP is a radiation summation over an area.

224 The data were grouped into two different satellite categories to train different RF models.
225 One RF model grouping for GOES geostationary satellites and one for other polar orbiting
226 satellites. The following satellites were used as FRP inputs in training the RF models: GOES for
227 the geostationary satellite; MODIS-Aqua and Terra , NOAA-20, S-NPP for the polar orbiting
228 satellites [Csiszar 2014, Giglio 2016]. These satellite inputs were chosen to remain consistent
229 with the current FRP ingested values for HRRR-Smoke modeling methods. Combining satellite
230 detected FRP from polar orbiters might present some sensor biases, specifically along the end of
231 swaths [Wang 2018 and Li 2018]. Depending on time of day, there might be higher or lower
232 uncertainty based on where the sensor detected the FRP in the model grid point. Additionally,
233 swath angles and corresponding atmospheric composition can impact FRP values leading to over
234 or under representation in FRP values. These issues are addressed by [Polivka 2016 and Zhou
235 2023] and could imply that there are double counted points in the same model grid cell. Lastly,

236 there are only so many passes through the day over the same point, which would skew model
237 performance based on time of day. Since these hourly aggregated FRP values are also the
238 targeted truth output, the model will carry some of these errors. Additionally there are biases in
239 the polar orbiting models because of retrieval time of day. Satellite derived FRP can be difficult
240 to measure. For instance, as a satellite captures FRP over an area, if there are thick clouds and/or
241 smoke plumes created by the fire, FRP directly under the satellite's line of vision can be missed.
242 One example of this is on the Colorado front-range in instances when mountain wave clouds can
243 block the FRP detection in GOES East but capture it in GOES West. These missed detections
244 can cause gaps in hourly or even a day's worth of FRP retrievals. It cannot be assumed an FRP
245 point from one satellite pixel will be measured in the next hourly FRP detection.

246 In addition to missing fire points, there are instances of false FRP detections either from
247 flares on objects caused by the sun, such as intense reflections on solar panels, or other false
248 detections that have nearly identical signatures to fire produced radiative power. These can occur
249 depending on the solar angle and satellite angle, time of year, etc. These are known issues and
250 difficult to account for when working with FRP data. For RF training, validation, and testing, a
251 condition was applied to each non-zero FRP RAP grid cell that the same RAP grid cell must
252 have a non-zero FRP from the day before to remove controlled burns and/or reduce false
253 detections. This might miss the first day of a fire or a short-lived fire, limiting the application to
254 multi-day fires, but in general was a good filter for emphasis on wildfire modeling and case
255 studies. Those models trained on western domains further target the wildfire application. This
256 study assumes that there are some of these false detections that might be included in the data,
257 especially with the GOES collected data.

258 Lastly, controlled wildfire burns produce FRP values and these fires are often very
259 predictable in behavior, short lived, and have weaker FRP values. They were not the intended
260 modeling goal of the project. The goal was to model longer-lived wildfires and large
261 aerosol/smoke contributors, so considering how to filter controlled burns out played into some of
262 the previously listed restrictions such as that of the yesterday mean FRP being greater than zero
263 on input data. By reducing controlled burns, the dataset might also have had a reduction in short-
264 instances of false FRP detection.

265

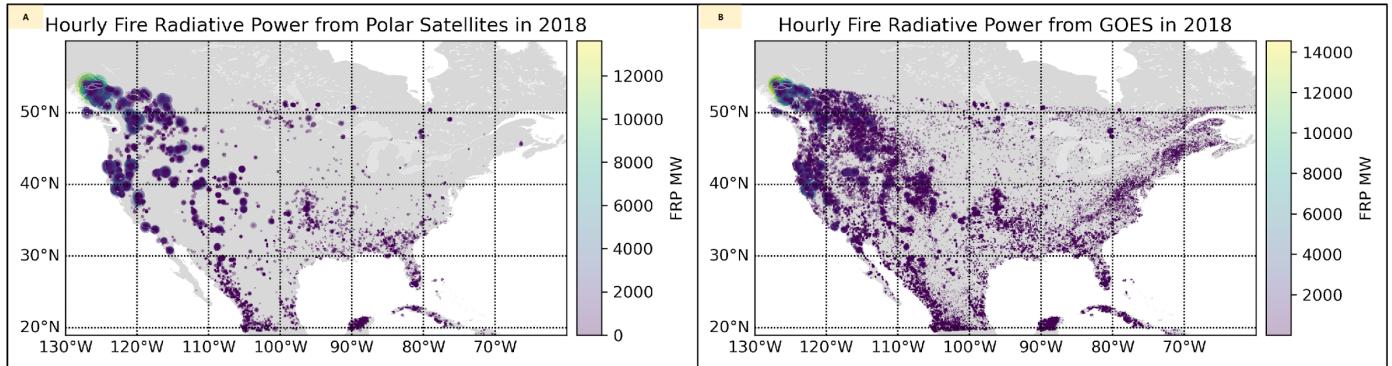
266 C) Calculating FRP on a gridded domain –

267 The RF models were designed for eventual use in numerical modeling on a gridded
268 domain. For consistency and application, the RF inputs and output are gridded to the RAP conus
269 domain with a resolution of roughly 13km. FRP is a product derived from satellite sensors and
270 data. Recall that depending on the sensor resolution, the magnitude of the FRP will be different.
271 Polar orbiter FRP values and magnitudes tend to be lower in value than FRP collected from a
272 geostationary satellite like GOES. To represent these FRPs on the RAP grid cell domain, the
273 polar orbiting satellite FRPs were aggregated over each grid cell for each hour of the day by their
274 latitude and longitude information stored with the FRP retrievals. The corresponding RAP grid
275 cell containing the FRP point was found by taking all centroid latitude and longitude from the
276 RAP model grid and doing a shortest distance calculation for each satellite FRP latitude and
277 longitude to find the cell containing that FRP's latitude and longitude. This was done for each
278 logged FRP value. Particularly for the polar orbiters, there would be several FRP measurements
279 that were located within the same model grid as another polar orbiting FRP point for that hour,
280 due to the high-resolution nature of these satellites. Because FRP is power, for all same-hour
281 satellite FRP that were contained in the single RAP grid cell, the FRP would be aggregated to
282 provide one FRP value for that hour and grid cell.

283 While GOES is lower resolution than the RAP, it is much lower resolution than polar
284 satellites and it was decided that for the geostationary satellites that if there were multiple points
285 per model grid cell, then an average was taken between the two or more points. Initially, GOES
286 satellites were aggregated in the same way as the polar orbiters, however the totaled magnitude
287 for RAP grid cell was noticeably higher value than the polar and HRRR-smoke FRP inputs. So,
288 it was found that taking an average provided more comparable gridded hourly FRP values. There
289 are also fewer instances when multiple GOES FRP points occur in the same RAP grid cell for an
290 hour.

291 Once the hourly aggregated or mean RAP-gridded FRP of the satellite FRP values were
292 calculated, an additional calculation was made and stored in a separate variable for the same
293 RAP grid cell for that current hour. This is an average of the FRP for the day-before, or a 24-
294 hour mean from the hourly FRPs of 0-23 UTC the day before in that same model grid cell. For
295 example, if the calculation was on August 5th at 3 UTC, the average would be from 0-23 UTC on
296 August 4th and stored to the August 5th 3 UTC grid cell in a new variable called FRP_yester
297 (FRP yesterday). This serves as the previously mentioned yesterday mean FRP input value for

298 the model. The day-before mean FRP was the selected input value to remain consistent with the
299 current method of using previously observed FRP in the hourly FRP modeling calculation.



300
301 Figure 2: Distribution of all satellite FRP (MW) points across the conus domain from GOES east and west in
302 A, and then the polar orbiters in B for all of 2018. The size of the dot and color relate to the magnitude of the
303 FRP of the fire and the brighter colors are denser in observed hourly FRP points.

304
305 Mapping FRP to the RAP domain shows good coverage over both CONUS and western,
306 west of -105 longitude, domains in 2018. The plot in Figure 2 shows a distribution of hourly FRP
307 points over the entire domain for the GOES satellites. While the individual hourly FRP points
308 were not used in the training of these RF models, these hourly FRP are the outputs that the model
309 is trying to predict using the inputted mean values. The trend for higher and denser FRPs to be
310 located in the western part of the conus is likely related to the wildfire behavior. These higher
311 FRP values are of interest to wildfire modeling and therefore the RF models are trained on a
312 CONUS and a western domain, producing different models depending on the desired region.
313 There is more coverage in the GOES FRP panel plot. This shows that even with certain
314 thresholds in place to avoid false detections and controlled fires, there is still the chance for non-
315 wildfire points in the training-testing-validation dataset. However, GOES can be more sensitive
316 to FRP detection over the entire conus for all 24 hours in the day, and this plot might reveal
317 additional training points otherwise missed by only using the polar orbiters.

318

319 **3. Machine Learning:**

320

321 *A) The Random Forest (RF)-*

322 RF models are a combination of tree predictors such that each tree depends on the values
323 of a random vector sampled independently and with the same distribution for all trees in the

324 forest [Breiman 2001]. RF models could be considered one of the more simplistic approaches,
325 since there are more logical processes going on “under the hood” and the decision-making tree-
326 like structure and it is easier to tease out information about independent variable input
327 importance to the model’s solution. As stated before, importance values are related to the RF
328 itself and not necessarily or directly to the physical relationship. These importance scores are
329 valuable when wanting to better interpret potential relationships between modeled FRP and
330 weather variables, but are not a direct statistical relationship. Lastly, RF models tend to be more
331 resilient to outliers because of the aggregation structure of the model [Srivastava 2021]. Because
332 of these reasons, the RF was a well-suited structure for this particular problem.

333 Initially, to determine which ML structure to proceed with for the project, there were
334 experiments with different simplified machine learning model structures using a dataset created
335 from meteorological inputs and 2018 FRP satellite data. Between the support vector machine
336 (SVM) model, neural networks (NN), and RF, it was found that the RF performed the best, with
337 neural networks having slightly poorer performance MAE and accuracy measurements. The
338 forest in these experiments contained 1000 decision trees, which was found to produce best
339 results in the final models. While more extensive tests could have been conducted, and more
340 hyperparameters tuned for both the RF and other models in the initial tests, there was
341 improvement on the current method in hourly FRP modeling. Additionally, the previously
342 mentioned benefits of using the RF on this particular problem factored into the RF selection. It is
343 a simply structured model, and performed competitively to the current hourly FRP method with
344 default RF model parameters. Future applications will include rigorous model optimization
345 processes, but with a small set of inputs and default settings, the models performed competitively
346 with the existing method. Lastly, NN and deep learning can be valuable tools for geospatial
347 problems, but like all ML methods can run the risk of overfitting due to geospatial and temporal
348 memorization, especially on smaller and less event diverse datasets. This dataset includes hourly
349 FRP measurements from wildfires from 2018 and it was a concern that this is not a large or
350 diverse enough dataset to use deep learning and would benefit from a point-by-point input
351 method instead of a geospatial input. For instance, there are tens to hundreds of wildfires that
352 produce non-zero FRP for several weeks in 2018, which consequently produces thousands of
353 FRP training points, but those points are related in space and time. If a simpler method performs

354 as well as more complex methods trained on a more limited datasets, it is good practice to start
355 with and use the “simpler” method.

356

357 *B) RF variable input importance analysis-*

358 As previously mentioned, the primary satellite FRP source for the RF models is the FRP
359 average from the UTC day before, labeled “FRP_yester”. This selection for the RF models was
360 made for two reasons. The first is that hourly FRP is not always available in real time for the
361 modeling process, since there is a lag in when this information is received. Of those received
362 points, sometimes the hourly value might not actually be a fire and therefore should not be
363 included in hourly FRP and subsequently smoke modeling. Additionally, polar orbiters tend to
364 miss consecutive hourly FRP measurements due to orbit, which means that training the RF
365 models from previous hour FRP would be useless on polar FRP data. By using the yesterday
366 FRP average, the RF models remain consistent with the inputs for the current HRRR smoke
367 modeling methods.

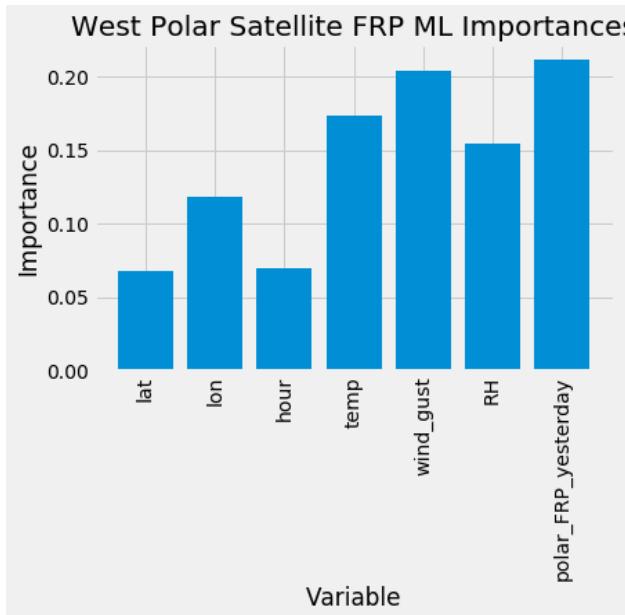
368 The second reason is the motivation to focus on wildfire-specific modeling. By using the
369 FRP_yester as an input with a requirement of training and testing on values greater than zero, it
370 is a good way to isolate modeling on fires that have persisted more than one day. There is
371 valuable information in the initial day of FRP data, however due to the data flow, these points
372 were not used in training this generation of RF hourly FRP models.

373 To determine the importance of each input variable and gain an understanding of what
374 the RF deemed to be the most important inputs in determining the hourly FRP, RF importance
375 calculations and plots were analyzed. The RF importance values can be determined using
376 python’s built in feature importance [scikit]. Input importance is a score assigned to each input in
377 the RF model and all scores total 1. Determining input importance is a fairly easy feature to use
378 when training RF models. It must be understood that the importance calculations and plots are
379 not directly indicating the physical importance or relationship of a variable to hourly FRP, but
380 rather these values indicate what the RF model determined most and least important in
381 calculating and modeling the hourly FRP. When testing initial inputs for the RF models,
382 additional variables, such as vegetation type and instantaneous precipitation were tested in an
383 initial RF model design but scored so lowly on the random forest importance variable diagram
384 that models trained without these inputs performed as well or better than the ones including the

385 inputs. The RF variable importance experiments for both polar and geostationary satellite trained
386 models showed that inputs like these could be removed from the input list without negative
387 impacts on the hourly FRP RF model performance. This could be due to the type of modeled
388 precipitation variable inputted. In future RF experiments, a different precipitation variable could
389 be tested. It could also be due to lag in FRP information and/or extremely rare event of
390 precipitation during an active fire compared to the rest of the dataset.

391 Figure 3 is an example of visualizing the importance of each input variable in the
392 determination of the final modeled hourly FRP for the polar trained RF on the western domain.
393 While not a direct physical relationship to FRP, much can be interpreted from these importance
394 plots about how the input variables relate to hourly FRP in the model. The plot shows that
395 previous knowledge of FRP is the most important variable in determining the future FRP value.
396 Next, wind, temperature, and relative humidity were the most important factors. The last three
397 inputs were considered least important, but not unimportant, and the model performed better
398 with these inputs included. The following inputs were used as final inputs in the final model
399 designs, and the output is that hour's FRP value for corresponding RAP grid point:

- 400
- 401 - Model Latitude Value
402 - Model Longitude Value
403 - Forecast Hour of day
404 - Yesterday mean FRP (MW)
405 - Forecast Hour 2m temperature (K)
406 - Forecast Hour Relative humidity (%)
407 - Forecast Hour 10m wind gust (m/s)
- 408



409

410 Figure 3: The plot shows an example of the distribution of RF determined importance rankings. This example
 411 is the breakdown of the importance scores for each input variable to the western CONUS domain polar-
 412 orbiting FRP RF model. The sum of the importances adds to 1. The polar_FRP_yesterday represents the day-
 413 before mean FRP for that corresponding RAP grid point.

414

415 There are additional fire fuel inputs that have been known to impact wildfire behavior,
 416 however due to the constraints of those variables temporal and spatial availability, they were not
 417 all viable choices to be used in the hourly FRP modeling pipeline. These models are designed to
 418 produce an hourly FRP value, which is a very short-term forecast. Using explainable ML tools
 419 like input importance revealed that not all initial selected inputs listed in section 2.A were
 420 necessary to train the best performing RF models. For instance, initial model training
 421 experiments used vegetation as an input, but when that variable was removed, the models
 422 showed either no change in performance or even performed slightly better. The importance value
 423 when compared to other inputs was nearly 0. This occurred with vegetation type and
 424 instantaneous precipitation and led to questions about why the model importance rankings might
 425 be that way.

426 Logically, precipitation should have a direct impact on a burning fire. One explanation as
 427 to why precipitation was found as a least-important variable and removed from the final RF
 428 models is the rareness in events where precipitation occurs during burning fires. Additionally,
 429 the model only uses the day before mean FRP as an input and perhaps the model therefore did

430 not find usefulness in hourly instantaneous precipitation values. There is a hypothesis that
431 precipitation could be more important to the modeled hourly FRP if either the previous hour FRP
432 or a rolling 24-hour average, could be used as an input. As for vegetation, one guess could be
433 that vegetation information, which doesn't change, is accounted for with location information.
434 Perhaps a more useful field for future experiment would involve fire fuel.

435

436 *C) Model Development Pipeline*

437 Once the individual satellite FRP points were aggregated to the corresponding RAP grid
438 cells, this data was merged with the initial hour weather data from that same-hour's RAP output.
439 The data were then stored into a csv table containing the hourly FRP, weather, location, and time
440 information from 2018. For each hourly FRP stored into the table, day-before mean is calculated
441 from 0-23 UTC by calculating an average of all the non-zero hourly FRP in that same RAP grid
442 cell. The RF models were trained from the tabular data, not gridded data. Each tabular point is a
443 unique gridded FRP for a unique hour in the year, meaning there are no redundant values for the
444 same grid cell at the same hour on the same. The data are divided randomly into 70% training
445 and 30% testing for the RF models. All RF models were trained with 1000 estimators and a
446 random state of 42. Analysis is completed on the testing data for 2018, and then case studies are
447 selected as blind validations of the trained RF models. Anything outside of 2018 was excluded
448 from RF model training.

449

450 **4. Results:**

451

452 *A) Random Forest Overall Performance-*

Satellite Input	GOES	Polar	GOES	Polar
Domain	Conus	Conus	West	West
Training Size	221887	38536	113870	26026
Testing Size	95095	16516	48802	11154
RF MAE	40.11	176.13	63.7	129.16
HRRR MAE	118.85	275.81	178.47	362.88
Variable Importance: lon	0.17	0.12	0.19	0.14
Variable Importance: lat	0.08	0.08	0.08	0.1
Variable Importance: RH	0.09	0.19	0.1	0.14
Variable Importance: temp	0.11	0.14	0.11	0.15
Variable Importance: hour	0.07	0.08	0.07	0.06
Variable Importance: wind_gust	0.12	0.2	0.12	0.19
Variable Importance: goes_FRP_yesterday	0.35	N/A	0.34	N/A
Variable Importance: polar_FRP_yesterday	N/A	0.19	N/A	0.22

453

454 Table 1: The table lists the primary four trained RF models, their domains, number of training and testing files.
 455 The MAE (MW) is calculated over the entire test set for both the current HRRR method and the corresponding
 456 RF for each hourly FRP satellite value where the corresponding yesterday mean FRP > 0. The importance
 457 scores for each input are also listed in this table and sum to 1.

458

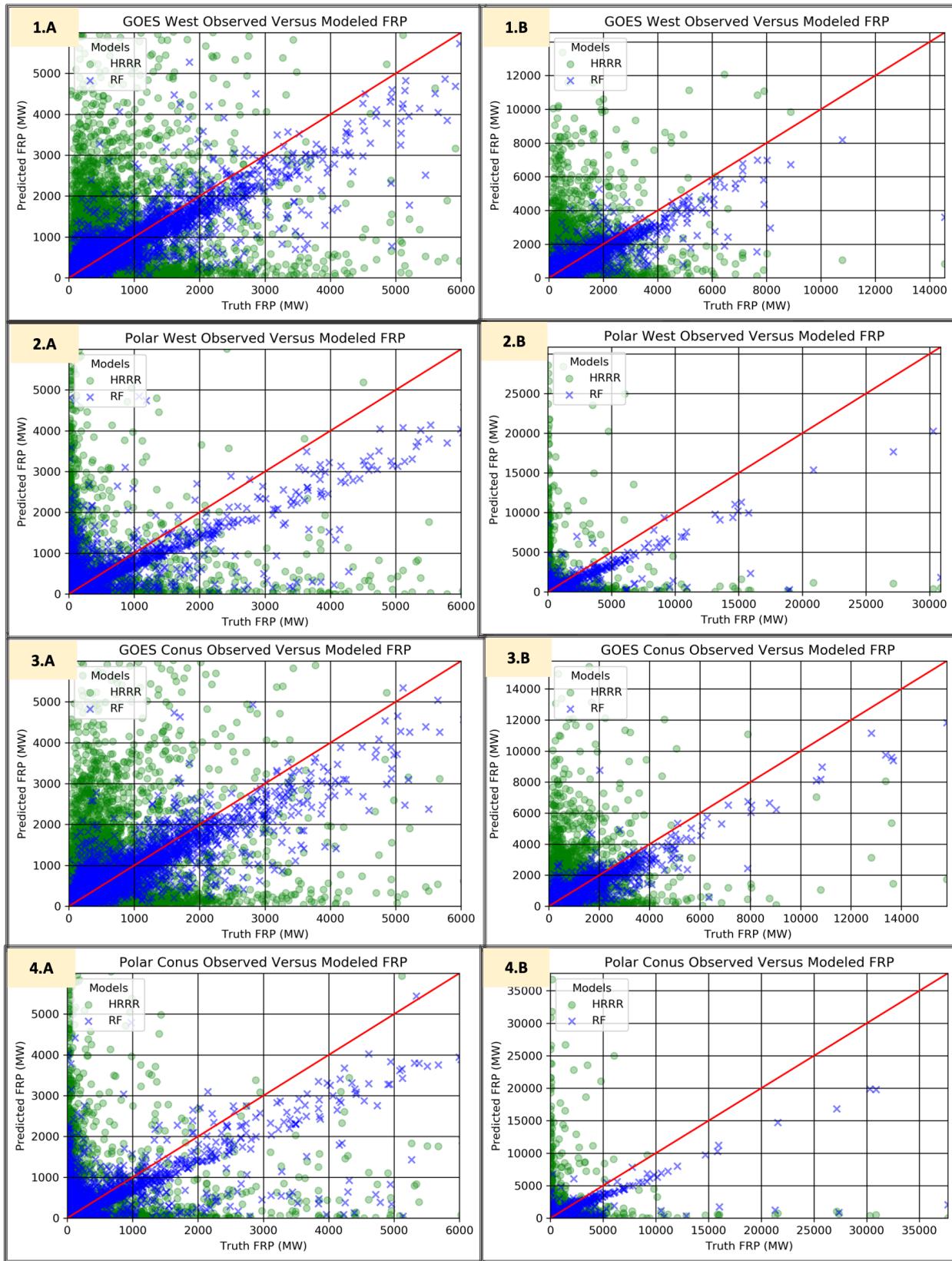
459 Table 1 shows the difference in training and testing sizes for the GOES and polar
 460 satellites, as well as between the two different domains. The western polar trained RF model had
 461 the fewest training and testing samples and also had the highest MAE over the 2018 testing
 462 dataset. This table also shows that in general, the yesterday mean FRP is the most important
 463 input to modeling the hourly FRP, but that the gap between that and other input variables is
 464 smaller for polar orbiters than GOES. Wind gust and longitude tend to be variable inputs of
 465 higher importance followed by relative humidity and temperature. For all four models, the RF
 466 model produced lower MAE than the current HRRR method. The CONUS-trained models had a
 467 better performance than the western domain trained models, likely due to the increased training
 468 quantity. Likewise, the GOES trained models had much lower MAE than the polar models.

469 These MAE's were calculated from the 2018 test set and the next section will show results from
470 specific wildfires in 2018 as well as those from 2019 and later.

471 The RF, by design, will produce outputs that are more constrained to the 1:1 predicted:
472 truth line than other ML models because it performs less extrapolation. In this sense, there are
473 few RF predicted major outliers seen in all the panel plots in Figure 4. Figure 4 shows the testing
474 dataset spread of RF hourly FRP predictions in addition to the Gaussian persistence method used
475 by the HRRR. The x-axis is labeled "truth" or the hourly gridded-FRP for that corresponding
476 satellite. The y-axis the model predicted hourly FRP value. Plots 1 and 3 compare the conus and
477 western-only domains of both models using the same day-before RAP-gridded mean FRP input.
478 Plots 2 and 4 are the same but using polar orbiting satellite input. Unlike the HRRR method, the
479 RF models are more closely bunched with their predicted values. This has modeling benefits,
480 such that the model is much less likely to produce any unexpected, unphysical, or "wild" FRP
481 outlier predictions. The B panels in Figure 4 show that the RF can predict more extreme hourly
482 FRP values, however they all have the tendency to under predict. But when compared to the
483 HRRR method, the general fit, especially at those high values, is still an overall improvement
484 with all four models.

485 All plots in figure 4 show that the RF models tend to under-predict more extreme FRP
486 values over the 2018 testing dataset and the HRRR model struggles to match these values as
487 well, either by over or under estimating the hourly FRP.

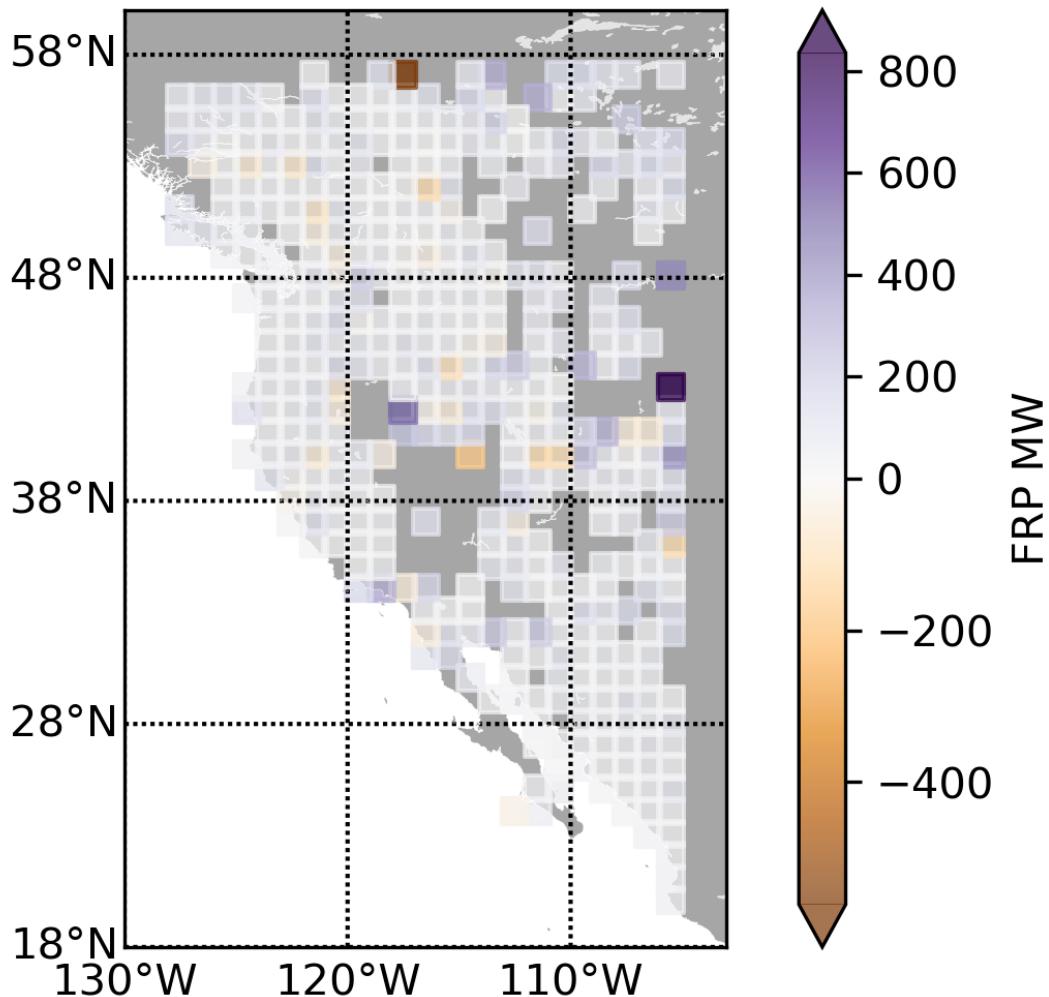
488



490 Figure 4: Scatter plots for different models compared to the observed satellite hourly FRP (MW) for the testing
491 hourly FRP points. The redline is the target 1:1 line. Blue crosses are the RF modeled hourly FRP and the
492 green dots are the corresponding HRRR modeled values. Panels 1-4 represent the four different trained RF
493 models with different satellite sources or trained regions. The A panels offer a close look at the lower-hourly
494 FRP observed and modeled hourly FRP. The B panels are bounded by the maximum observed hourly FRP.
495

496 The models were not trained with specific physical constraints and nothing was explicitly
497 programmed into the model. There were initial restrictions on which data could be used for
498 training and running the model; as stated before with the day-before mean FRP greater than 0
499 indicating a likely fire in that RAP grid cell the day before. This was done to better filter out
500 single-day fires and false detections. Single-day fires can often include controlled agricultural
501 burns, which have very predictable behaviors where the FRP might be higher at the beginning of
502 the burn but tapers to zero within a few hours due to its controlled nature. This means that after a
503 few hours, the modeled value would be trained to taper to zero, which doesn't happen often in
504 wildfire scenarios. While still important, controlled burns were not the focus of this modeling
505 goal and created an imbalanced fire-type dataset where wildfire hourly FRPs were outnumbered
506 greatly by controlled burn hourly FRPs. These controlled burns were deliberately reduced with
507 the data restrictions (or conditionals), as were false detections. Controlled burns were further
508 reduced when training the western domain specific models. While reduced, it is likely that there
509 is a small number of false detections or controlled fire FRP points still included in the final
510 training and testing datasets.

Averaged Error Between RF and Polar 2018



511

512 Figure 5: Above is a scatter plot over the western CONUS domain made from the western polar RF model
513 hourly FRP. Mean Error calculated by subtracting the observed polar hourly FRP from the modeled RF hourly
514 FRP at each observed hourly FRP point, and then taking the average of all those differences over 2018. Each
515 point is the average of all Mean Error values within the 1x1 degree grid points.

516

517 For analysis on general model behavior over the western domain for 2018, Figure 5 was
518 created by taking the mean of all average errors calculated within a 1-degree by 1-degree
519 bounding box around the RAP grid cells within the domain. Each hourly observed polar FRP
520 was subtracted from the same hour's RF modeled FRP. Then an average over all these
521 differences was taken for 2018 to represent one averaged error value for that area. Negative
522 values represent under-modeled hourly FRP, positive mean the RF over predicted. Over the

523 western domain for the polar RF model, the RF model's hourly predicted FRP tends to be just
524 slightly under or over predicted from the observed hourly FRP. This is consistent with Figure 4
525 plots showing the majority of RF dots clustered at lower values and either under or over the 1:1
526 line. When all points' errors are averaged over the year, it's difficult to extrapolate why certain
527 behavior might be happening. Figure 4 shows that the RF model under-estimates FRP at more
528 extreme values, however if the FRP over the same fire area is averaged over time, the model
529 generally overestimates FRP. One hypothesis as to why is that in using a yesterday mean FRP as
530 input to the RF models, there is already a missing knowledge of some of the more extreme FRP
531 values from the day before since they get smoothed out in the mean as well as the lag in
532 information.

533 This behavior was generally consistent between both satellite types and domains. Figure
534 5 shows that there is a slight trend for the polar western RF to overestimate FRP values at lower
535 values. The GOES models tended to have better performance and that could be due to a few
536 factors. The first is that GOES satellites take more temporally frequent hourly FRP
537 measurements, so there are more training points per fire as well as a more uniform distribution of
538 measurements throughout 24 hours. This can be seen in the scatter plots of Figure 2, where the
539 polar FRP vs. hour has a less uniform distribution from GOES FRP vs. hour. Polar orbiters
540 collect high resolution FRP measurements, but those points have been gridded to the RAP
541 domain, so the advantage of having more points over a fire is not as relevant for this modeling
542 problem. This means that there is one RAP grid FRP value for the polar and GOES satellites,
543 even if sensor-wise the polar orbiters gathered many more points over the same area for a
544 particular hour. Another factor would be related to the first in that polar orbiters might have more
545 blocked measurements from clouds or smoke plumes. Because of that and the orbits of the polar
546 satellites, GOES will naturally collect more frequent FRP measurements in a day without the risk
547 of points being blocked with smoke or clouds over that hour that the polar passed over the RAP
548 grid. If a satellite collects 100 points and misses 5, it is less an issue than a satellite that collects
549 25 points and misses 5.

550

551

552 *B) Case Studies-*

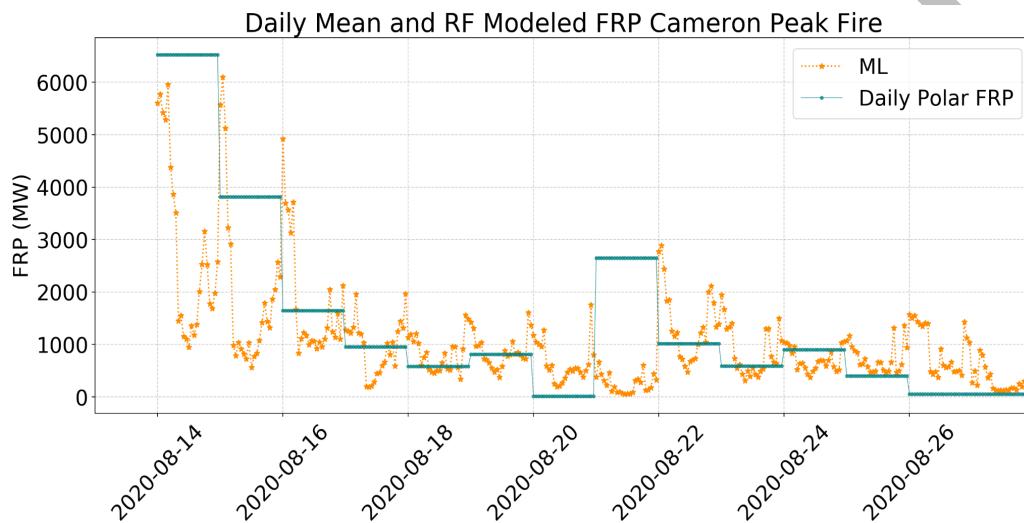
553 This section hand-selects western wildfire case studies from 2018-2021 to analyze
 554 performance of the RF against the existing method. All wildfires were in the western domain and
 555 corresponding western domain RF models were used to produce the hourly FRP RF modeled
 556 results. Fires in 2018 were included in the testing/training process of the RF model while fires in
 557 2019 and 2020 are data that were never seen by the RF model. Table 2 is a collection of a few of
 558 the wildfires examined and indicates which satellite RF model was used and the corresponding
 559 overall MAE for the RF and HRRR hourly FRP forecasting. The table shows that in all fire cases
 560 except for the Mendocino Complex fire, the RF outperforms the baseline HRRR method in
 561 MAE. One important note is that when available, the MAE was calculated from each satellite
 562 observed hourly FRP over the specified timeframe. Otherwise, the MAE is calculated from that
 563 day's mean FRP value. The last column indicates which calculation was performed for that fire.
 564

Fire Name	Where (Lat, Lon)	Dates	RF Model Type	RF MAE	HRRR MAE	MAE Calculated from hourly or mean FRP?
Cameron Peak	(40.609, -105.879)	8-13-2020 to 10-1-2020	GOES Western	2360.87	3565.99	mean
Cameron Peak	(40.609, -105.879)	8-13-2020 to 10-1-2020	Polar Western	2551.46	4187.60	mean
Camp	(39.810, -121.437)	11-7-2018 to 11-15 2018	Polar Western	2268.44	8246.00	hourly
Creek	(37.192, -119.261)	9-5-2020 to 11-7-2020	GOES Western	3500.00	4737.94	mean
Creek	(37.192, -119.261)	9-5-2020 to 11-7-2020	Polar Western	3235.13	4737.94	mean
Lionshead	(44.723, -121.678)	8-7-2021 to 11-12-2021	Polar Western	1447.29	2334.18	mean
Mendocino Complex	(39.24, -123.1)	7-26-2018 to 8-29-2018	Polar Western	848.85	575.72	hourly
Pine Gulch	(39.336, -108.526)	8-1-2020 to 8-13-2020	GOES Western	186.16	285.53	mean
Watson Creek	(42.5, -120.7)	8-14-2018 to 8-31-2018	Polar Western	1396.47	1609.42	hourly
Williams Flat	(47.98, -118.624)	8-1-2019 to 8-9-2019	GOES Western	467.22	870.34	mean
Williams Flat	(47.98, -118.624)	8-1-2019 to 8-9-2019	GOES Western	429.46	743.50	hourly

565 Table 2: The table above lists several western wildfires spanning the 2018-2021 fire seasons. The names, dates,
 566 and locations are provided in the table. The tested RF model is also listed in the table where the GOES
 567 Western indicates the RF trained from GOES satellite FRP points over the western conus and the Polar
 568 Western indicates the RF trained from polar orbiters over the western conus. The MAEs (MW) were calculated
 569 over the entire time duration of the fire between observed FRP and the RF or HRRR model methods. The last
 570 column indicates if the MAEs used in the total MAE were calculated at each hour from the true daily mean of
 571

572 the hourly FRP of that day (the absolute value of the modeled hourly FRP minus daily observed mean FRP) or
573 if it was calculated from the hourly FRP points (the absolute value of the modeled hourly FRP minus same
574 hour's observed FRP). It should be noted that the fires in 2018 were included in the training-testing dataset.
575

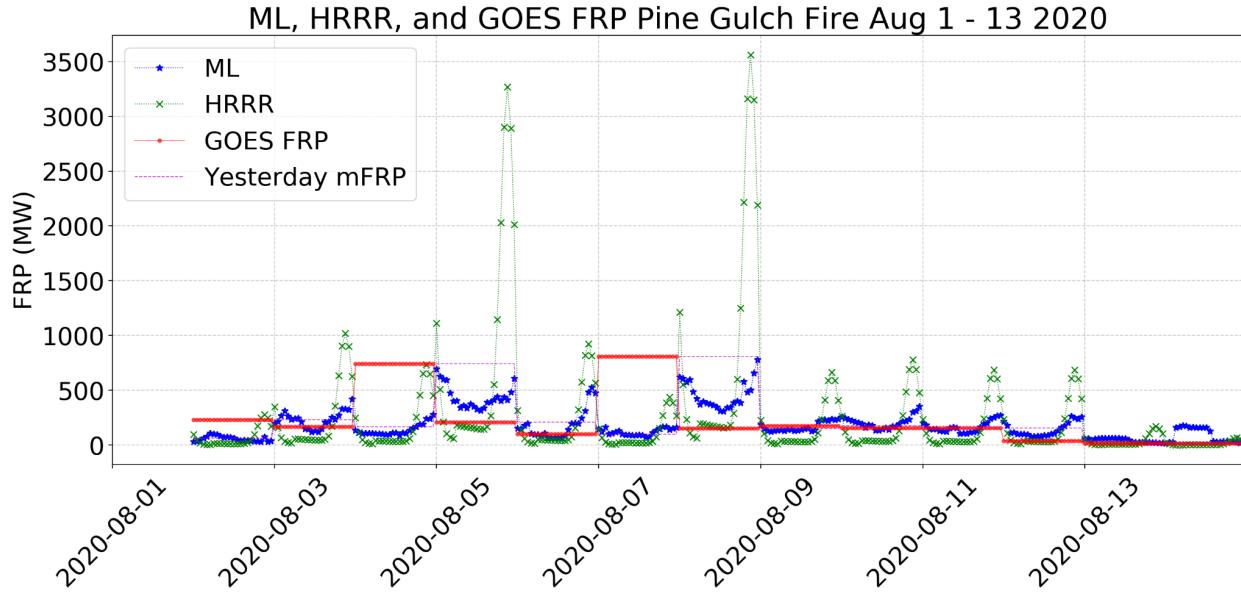
576 Some fires have shorter durations and others, such as the Cameron Peak fire, have
577 periods of low FRP and then high flare ups. This makes the wildfire behavior and hourly FRP
578 harder to predict, especially with the input lag of yesterday's FRP. However, there is substantial
579 improvement in the RF predicted hourly FRP than the HRRR method for the Cameron Peak fire
580 in both the polar and GOES RF models.



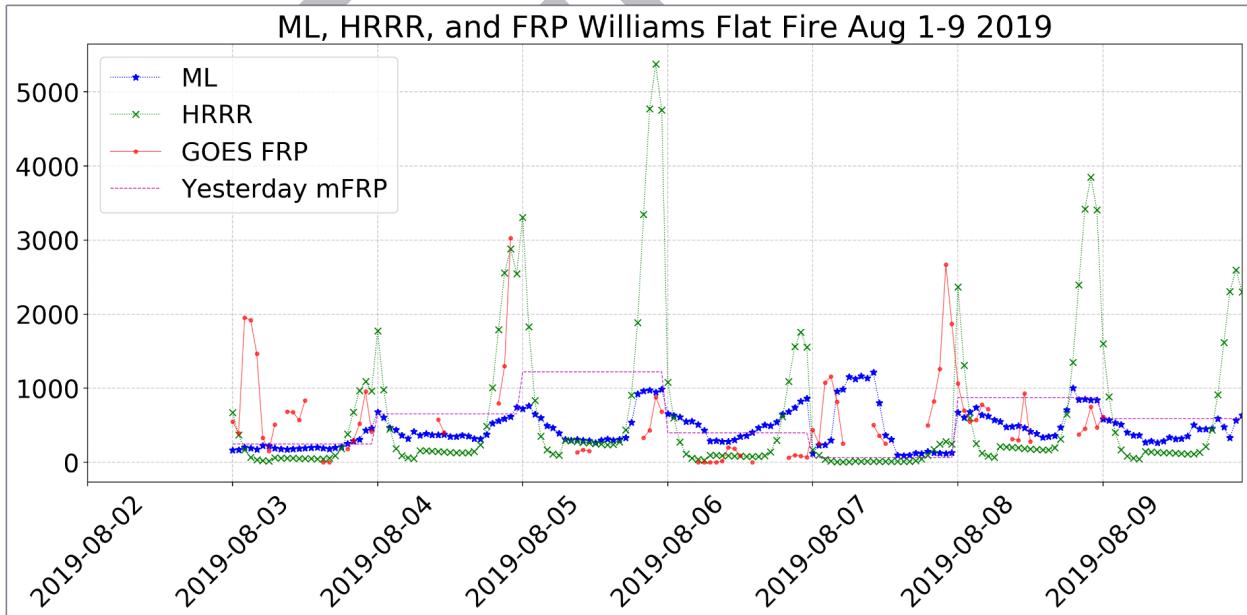
581
582 Figure 6: Plotted is a section of the Cameron Peak Fire where the orange is the RF modeled hourly FRP (MW)
583 and the blue is the current-day mean FRP.

584
585 We observed the RF models all had Gaussian-like trends, without any explicit knowledge
586 of the persistence Gaussian model. Remember that the RF and HRRR method both use a single
587 day-before mean as input, which does not change over the iterations of hours the model is run the
588 next day. This means that the models do not get updated FRP information until the following day
589 when a new mean is calculated. The RF models likely learned the hourly relationship between
590 FRP and hour of day while training and testing. Unlike HRRR the persistence method, however,
591 the RF model makes adjustments based on all model inputs, and doesn't follow a strict curve.
592 Each curve is also unique based on the fire or RAP grid that is being modeled. One way to
593 analyze RF performance is in Figure 6, which shows a subset of the Cameron Peak fire's daily
594 mean polar FRP compared to the hourly RF modeled FRP, given in UTC. The conclusion from

595 this figure is that in general, the modeled hourly FRP doesn't exceed the daily mean FRP. This is
 596 not always the case, but a good example of the typical RF behavior over the span of a day
 597 compared to the mean FRP of that same day for a fire.



598
 599 Figure 7: This is a time series of the Pine gulch fire where mean daily FRP (MW) is in red,
 600 mean yesterday FRP is in the dashed purple (note the day shift), and the two models are in the hourly points in
 601 green (HRRR-Smoke) and blue (RF model).
 602



603
 604 Figure 8: This is a time series of the William Flat fire where the GOES hourly observed FRP (MW) is in red,
 605 the input mean yesterday FRP is in the dashed purple (representing the mean of the hourly FRP the day

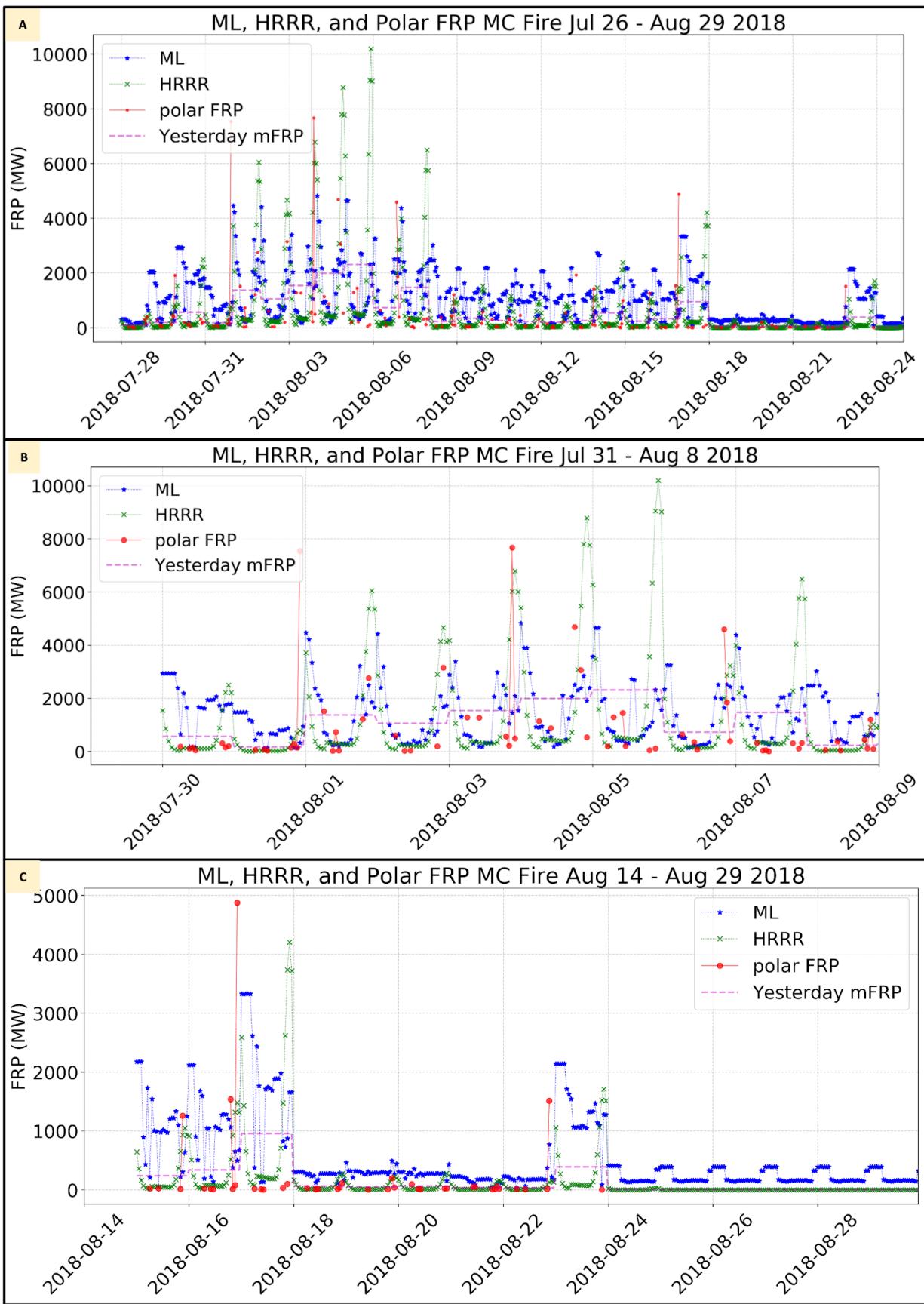
606 before), and the two models are in the hourly points in green (HRRR) and blue (RF model). The missing
607 hourly FRP points are due to missing measurements from the satellite over the fire.

608

609 This Gaussian mimic behavior by the RF models is also noted in Figures 7 and 8 for the
610 Pine Gulch and Williams Flat fires. This figure compares the daily mean FRP to the individual
611 hourly modeled FRP, illustrating an example how RF model behaves differently than the
612 persistence in hourly FRP forecasts. In Figure 7, the Pine Gulch fire daily mean FRP and
613 yesterday mean FRP are plotted in red and purple, respectively. Recall that the purple yesterday,
614 or day-before, mean FRP is used as the input for the current day's hourly FRP modeling. As is
615 consistent before, the general behavior of the RF model is to be more conservative in hourly FRP
616 values as well as fit similar magnitudes to the mean FRP unlike the HRRR persistence method,
617 which can be seen to model two distinct high maximums in hourly FRP. On August 5th, the truth
618 daily mean FRP is lower than the day before, but there is a noticeable spike near the end of the
619 UTC day in RF and HRRR modeled FRP, likely due to the day-before input. Unlike the
620 persistence, the RF model stays more conservative in its spike whereas HRRR persistence's
621 spike is over 2,500 more in magnitude than the daily mean. However, the persistence midday
622 modeled FRP is closer to the daily mean. By analyzing case studies against the daily mean shows
623 that both modeling methods provide more insight in 24-hour FRP behavior and further
624 verification is needed to determine which of the two characteristics might be the better or more
625 tolerated in practice. Another benefit to introducing meteorological and location information into
626 a RF model is that there are hourly deviations in the overall curve for the day as opposed to the
627 strictly Gaussian shape of the HRRR persistence model.

628 The Williams Flat fire in Figure 9 provides a new perspective on analyzing the modeled
629 hourly RAP-gridded FRP to the truth RAP grided hourly FRP, when available. The observed
630 RAP-gridded hourly FRP (truth) in red dots, the yesterday mean FRP value used as model input
631 in faint dashed purple, the HRRR method in green crosses, and the ML method in blue stars. The
632 yesterday mean is included in plots to reference the input FRP that the model receives for that
633 day's hourly FRP forecast. GOES rap-gridded hourly values are connected by line if there are
634 continuous hour observations. By looking at the modeled hourly FRP compared to the truth
635 hourly FRP, further insight can be gained in the fire-specific FRP over a 24-hour and longer
636 span. By keeping the day-before mean in the plot, the input into that day's hourly modeled FRP

637 can be easily referenced. In Figure 8, there are three distinct times it looks like the persistence
638 model captures the spike in hourly FRP better than the RF model on Aug 3rd, 4th, and 7th.
639 However, on August 5th, the persistence model has a large miss in the later hourly FRP modeling
640 while the RF does a much better at matching the hourly FRP observations as well as a much
641 better job modeling the afternoon FRP. These large spikes in the HRRR method and large misses
642 in FRP value contribute the most to the HRRR's MAE score. On August 7th in the earlier UTC
643 hours, the RF model does a very good job at modeling the observed hourly FRP, even when the
644 model input yesterday mean FRP was so low. Unfortunately, in the later UTC hours, the model
645 misses the observed FRP by large quantities while the HRRR model then performs better. Again,
646 on August 8th, the RF model seems to match the observed hourly FRP whereas the HRRR model
647 underestimates in the early UTC hours and then substantially overestimates in the afternoon
648 hours. Both data from the Pine Gulch as well as Williams Flat fire were not used in training and
649 testing the RF model and are completely blind RF tests. The models have noticeable differences
650 in the hourly modeled FRP and it can be concluded that the RF model can predict more unique
651 hourly FRP forecasts, sometimes better and/or within a closer magnitude than the persistence
652 method, but it is unclear which model is better at predicting hourly FRP in wildfire application.
653 The RF models are a good method to provide unique hourly FRP points that deviate from the
654 Gaussian curve structure and are not largely under or over estimating hourly FRP more than the
655 current method.



657 Figure 9: This is a time series of the Mendocino Complex Fire from July 26 – August 29 2018. Panel A is a
658 fuller time and includes the later flare-up. Panel B is a subset of July 30 - August 8 2018. Panel C is a subset of
659 the flare up of the fire August 14-29 2018. The polar hourly observed FRP (MW) is in red, the input mean
660 yesterday FRP is in the dashed purple (representing the mean of the hourly FRP the day before), and the two
661 models are in the hourly points in green (HRRR) and blue (RF model). The missing hourly FRP points are due
662 to missing measurements from the satellite over the fire.

663

664 For the Mendocino Complex case studies from 2018, recall some of the data were part of
665 the training and some of the testing dataset, and this feature is shown by the model learning
666 about the hourly FRP spikes is reflected in the plot when the RF performs well on FRP spikes.
667 However, this fire is included because the fire had a long period of low activity before spiking
668 again in the future and it was worth visualizing. The panels in Figure 9 show that RF does a good
669 job at capturing some of those extreme hourly FRP jumps, especially in the first third of the fire
670 duration. Additionally, the RF modeled hourly FRP tends to fluctuate well above the mean FRP
671 value, especially near the end of the timeline. While more days are shown in this plot, there still
672 is indication of the Gaussian behavior in both the observed polar hourly FRP as well as the RF
673 model. This is the only case study where the MAE for HRRR was lower than the RF, which is
674 interesting since points in this fire were used in training and testing the RF model. This is another
675 reason deciding which model version is overall better is difficult, especially near the end of the
676 time series when the RF over-estimates the observed hourly FRP for this fire.

677

678 **5. Conclusion:**

679 There are many motivations to develop machine learning models to predict hourly FRP
680 short term. Table 1 shows the results from four different RF models, which were developed and
681 analyzed to show that RF models can be used to model weather-specific hourly FRP from either
682 GOES or polar orbiting satellites and overall improved the performance compared to the current
683 HRRR method for modeling hourly FRP. These RF models had an overall lower MAE than the
684 existing HRRR Gaussian persistence model. The RF models also had overall lower MAE for
685 selected fire case studies in all but one instance.

686 The RF models tend to underestimate high extremes in hourly FRP values, potentially
687 due to the lag in information received by the RF model. The RF models generally follow the
688 daily mean FRP well. One bias in the dataset could be that there are far fewer points of very high

689 hourly FRP than lower FRP values, and these are further reduced by using a yesterday daily
690 mean as input. Since these high FRP events are rare, but of high interest, it is important to
691 improve current RF models to be more representative of those instances. The RF models do not
692 produce extreme hourly FRP values in modeled hourly FRP, unlike the HRRR method, and that
693 typically reflects better RF performance for fire FRP in the corresponding hours of day. The
694 behavior of the HRRR method for modeling high values of hourly FRP often produces an over-
695 estimated FRP, and therefore contributes to higher MAEs. However, when spikes do happen in a
696 fire, then that model can have lower errors in hourly FRP than the RF models. When compared
697 over an entire 2018 training and testing year as well as specific non-2018 fire case studies, the
698 RF models perform as well and better than the current modeling method and also produce unique
699 hourly values for each input based on weather inputs and additional information, as opposed to
700 being modeled directly to the Gaussian curve. It is difficult to determine exactly if the RF models
701 are better, but evidence in the case studies and overall MAEs show that RF models are able to be
702 considered an alternative to the standard persistence method. They provide a different behavior
703 of modeling hourly FRP and the extreme modeled hourly FRP spikes do not appear to occur any
704 more than the current method as can be seen in the scatter analysis in Figure 4 and the case study
705 figures. Further verification needs to be done to test the impact differences between the two
706 models in applications like smoke modeling.

707 This work is being continued with development of RF models where the training is over a
708 larger domain, timeframe, and test set. These models will include new input weather variables as
709 well as a new FRP satellite product.

710

711 *In dedication to those neighbors and friends who lost pets, homes, memories, and were displaced*
712 *by the Marshall Mesa fire December 30th, 2021.*

713

714 *Acknowledgments:*

715 *This research was supported (in part / by / etc.) NOAA cooperative agreement*
716 *NA22OAR4320151. This research was further supported (in part / by / etc.) NOAA cooperative*
717 *agreement NA22OAR4320151, for the Cooperative Institute for Earth System Research and Data*
718 *Science (CIESRDS). We would also like to acknowledge Ivan Csiszar, Johana Romero-Alvarez,*
719 *and Marina Tsidulko for their assistance and expertise involving the satellite FRP data.*

720
721

722 **Data Availability:**

723

724 Data analyzed in this study were a re-analysis of existing data, which are openly available at
725 locations cited in the reference section. The processed tabular data used in training and testing
726 the random forests can be available upon request by NOAA GSL.

727

728 **References:**

- 729 Ager, A. A., Day, M. A., Alcasena, F. J., Evers, C. R., Short, K. C., & Grenfell, I. (2021).
730 Predicting paradise: Modeling future wildfire disasters in the western US. *Science of The Total
731 Environment*, 784, 147057. doi:10.1016/j.scitotenv.2021.147057
- 732 Ahmadov R, Grell G, James E, Csiszar I, Tsidulko M, Pierce B, et al. Using VIIRS Fire Radiative
733 Power data to simulate biomass burning emissions, plume rise and smoke transport in a real-time
734 air quality modeling system. 2017 IEEE International Geoscience and Remote Sensing
735 Symposium. IEEE International Symposium on Geoscience and Remote Sensing IGARSS. New
736 York: IEEE; 2017. p. 2806-8.
- 737
- 738 Andela N, Kaiser JW, van der Werf GR, Wooster MJ. New fire diurnal cycle characterizations to
739 improve fire radiative energy assessments made from MODIS observations. *Atmos Chem Phys*.
740 2015;15(15):8831-46.
- 741 Bakhshaii, A., & Johnson, E. A. (2019). A review of a new generation of wildfire–Atmosphere
742 modeling. *Canadian Journal of Forest Research*, 49(6), 565–574. doi:10.1139/cjfr-2018-0138
- 743 Benjamin S., Grell G., Brown J., Smirnova T., Bleck R. (2004): Mesoscale weather prediction
744 with the RUC hybrid isentropic - terrain-following coordinate model. *Monthly Weather Review*,
745 132(2), 473-494. doi:10.1175/1520-0493(2004)132<0473:MWPWTR>2.0.CO;2

- 746 Benjamin S., Smirnova T., James E., Lin L.-F., Hu M., Turner D., He S. (2022). Land-snow data
747 assimilation including a moderately coupled initialization method applied to NWP. *Journal of*
748 *Hydrometeorology*, 23(6), 825-845. doi:10.1175/JHM-D-21-0198.1
- 749 Benjamin, S., James E., Hu M., Alexander C., Ladwig T., Brown J., Weygandt S., Turner D.,
750 Minnis P., Smith W. Jr., Heidinger A. (2021). Stratiform cloud-hydrometeor assimilation for
751 HRRR and RAP model short-range weather prediction. *Monthly Weather Review*, 149(8), 2673-
752 2694. doi:10.1175/MWR-D-20-0319.1
- 753 Benjamin S., Weygandt S., Brown J., Hu M., Alexander C., Smirnova T., Olson J., James E.,
754 Dowell D., Grell G., Lin H., Peckham S., Smith T.-L., Moninger W., Kenyon J., Manikin G.
755 (2016). A North American hourly assimilation and model forecast cycle: The Rapid Refresh.
756 *Monthly Weather Review*, 144(4), 1669-1694. doi:10.1175/MWR-D-15-0242.1
- 757 Breiman, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the
758 author). *Statistical Science*, 16(3). doi:10.1214/ss/1009213726
- 759 Csiszar I, Schroeder W, Giglio L, Ellicott E, Vadrevu KP, Justice CO, et al. Active fires from the
760 Suomi NPP Visible Infrared Imaging Radiometer Suite: Product status and first evaluation results.
761 *J Geophys Res-Atmos.* 2014;119(2):803-16.
- 762 DENNISON, P., CHAROENSIRI, K., ROBERTS, D., PETERSON, S., & GREEN, R. (2006).
763 Wildfire temperature and land cover modeling using Hyperspectral Data. *Remote Sensing of*
764 *Environment*, 100(2), 212–222. doi:10.1016/j.rse.2005.10.007
- 765 Dennison, P. E., Brewer, S. C., Arnold, J. D., & Moritz, M. A. (2014). Large wildfire trends in the
766 Western United States, 1984-2011. *Geophysical Research Letters*, 41(8), 2928–2933.
767 doi:10.1002/2014gl059576
- 768 Feature importances with a forest of trees. (n.d.). *scikit*. https://scikit-learn.org/stable/auto_examples/ensemble/plot_forest_importances.html. Accessed 19 July 2023
- 770 Giglio L, Schroeder W, Justice CO. The collection 6 MODIS active fire detection algorithm and
771 fire products. *Remote Sens Environ.* 2016;178:31-41.

- 772 Gu, J., Lu, Z., Li, H., & Li, V. O. K. (2016, June 8). Incorporating copying mechanism in sequence-
773 to-sequence learning. *arXiv.org*. <https://arxiv.org/abs/1603.06393>. Accessed 24 June 2022
- 774 Jain, P., Coogan, S. C. P., Subramanian, S. G., Crowley, M., Taylor, S., & Flannigan, M. D. (2020).
775 A review of Machine Learning Applications in wildfire science and management. *Environmental
776 Reviews*, 28(4), 478–505. doi:10.1139/er-2020-0019
- 777 James E., Benjamin S., Jamison B. (2020). Commercial-aircraft-based observations for NWP:
778 Global coverage, data impacts, and COVID-19. *Journal of Applied Meteorology and Climatology*,
779 59(11), 1809-1825. doi:10.1175/JAMC-D-20-0010.1
- 780 Joseph, M. B., Rossi, M. W., Mietkiewicz, N. P., Mahood, A. L., Cattau, M. E., St. Denis, L. A.,
781 et al. (2019). Spatiotemporal prediction of wildfire size extremes with bayesian finite sample
782 maxima. *Ecological Applications*, 29(6). doi:10.1002/eap.1898
- 783 Li, F., Zhang, X., Kondragunta, S., & Csiszar, I. (2018). Comparison of fire radiative power
784 estimates from viirs and Modis Observations. *Journal of Geophysical Research: Atmospheres*,
785 123(9), 4545–4563. doi:10.1029/2017jd027823
- 786 Linn, R., Reisner, J., Colman, J. J., & Winterkamp, J. (2002). Studying wildfire behavior using
787 FIRETEC. *International Journal of Wildland Fire*, 11(4), 233. doi:10.1071/wf02007
- 788 LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.
789 doi:10.1038/nature14539
- 790 Mamalakis, A., Ebert-Uphoff, I., & Barnes, E. A. (2022). Neural network attribution methods for
791 problems in geoscience: A novel synthetic benchmark dataset. *Environmental Data Science*, 1.
792 doi:10.1017/eds.2022.7
- 793 McGovern, A., Lagerquist, R., John Gagne, D., Jergensen, G. E., Elmore, K. L., Homeyer, C. R.,
794 & Smith, T. (2019). Making the black box more transparent: Understanding the physical
795 implications of machine learning. *Bulletin of the American Meteorological Society*, 100(11),
796 2175–2199. doi:10.1175/bams-d-18-0195.1

- 797 Oliveira, S., Rocha, J., & Sá, A. (2021). Wildfire risk modeling. *Current Opinion in Environmental
798 Science & Health*, 23, 100274. doi:10.1016/j.coesh.2021.100274
- 799 Polivka, T. N., Wang, J., Ellison, L. T., Hyer, E. J., & Ichoku, C. M. (2016). Improving nocturnal
800 fire detection with the VIIRS day-night band. *IEEE Transactions on Geoscience and Remote
801 Sensing*, 54(9), 5503-5519.
- 802 Roberts, G., Wooster, M. J., Perry, G. L., Drake, N., Rebelo, L.-M., & Dipotso, F. (2005). Retrieval
803 of biomass combustion rates and totals from fire radiative power observations: Application to
804 southern Africa using geostationary SEVIRI imagery. *Journal of Geophysical Research*,
805 110(D21). doi:10.1029/2005jd006018
- 806 Samudrala, S. (2018). *Machine Intelligence Demystifying Machine Learning, neural networks and
807 deep learning*. Chennai: Notion Press.
- 808 Scott, I. A. (2021). Demystifying machine learning: A Primer for Physicians. *Internal Medicine
809 Journal*, 51(9), 1388–1400. doi:10.1111/imj.15200
- 810 Srivastava, A. (2021, April 19). Let's talk about random forests! *Medium*. Analytics Vidhya.
811 <https://medium.com/analytics-vidhya/lets-talk-about-random-forests-524ae1138d8b#:~:text=Random%20forests%20are%20robust%20to,output%20of%20multiple%20decision%20trees>. Accessed 24 June 2022
- 814 Turco, M., Rosa-Cánovas, J. J., Bedia, J., Jerez, S., Montávez, J. P., Llasat, M. C., & Provenzale,
815 A. (2018). Exacerbated fires in Mediterranean Europe due to anthropogenic warming projected
816 with non-stationary climate-fire models. *Nature Communications*, 9(1). doi:10.1038/s41467-018-
817 06358-z
- 818 Vermote, E., Ellicott, E., Dubovik, O., Lapyonok, T., Chin, M., Giglio, L., & Roberts, G. J. (2009).
819 An approach to estimate global biomass burning emissions of organic and black carbon from
820 Modis fire radiative power. *Journal of Geophysical Research*, 114(D18).
821 doi:10.1029/2008jd011188
- 822 Wang, J., Yue, Y., Wang, Y., Ichoku, C., Ellison, L., & Zeng, J. (2018). Mitigating satellite-
823 based fire sampling limitations in deriving biomass burning emission rates: Application to WRF-

- 824 chem model over the Northern Sub-Saharan African region. *Journal of Geophysical Research: Atmospheres*, 123(1), 507-528.
- 825
- 826 Westerling, A. L. R. (2016). Increasing western US forest wildfire activity: Sensitivity to
827 changes in the timing of Spring. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1696), 20150178. doi:10.1098/rstb.2015.0178
- 827
- 828
- 829 Wiedinmyer, C., Akagi, S. K., Yokelson, R. J., Emmons, L. K., Al-Saadi, J. A., Orlando, J. J.,
830 and Soja, A. J.: The Fire INventory from NCAR (FINN): a high resolution global model to
831 estimate the emissions from open burning, *Geosci. Model Dev.*, 4, 625–641,
832 <https://doi.org/10.5194/gmd-4-625-2011>, 2011
- 833
- 834 Wiggins, E. B., Soja, A. J., Gargulinski, E., Halliday, H. S., Pierce, R. B., Schmidt, C. C., et al.
835 (2020). High temporal resolution satellite observations of fire radiative power reveal link
836 between fire behavior and aerosol and gas emissions. *Geophysical Research Letters*, 47(23).
837 doi:10.1029/2020gl090707
- 838
- 839 Wooster, M. J., Roberts, G., Perry, G. L., & Kaufman, Y. J. (2005). Retrieval of biomass
840 combustion rates and totals from fire radiative power observations: FRP derivation and
calibration relationships between biomass consumption and fire radiative energy release. *Journal
of Geophysical Research*, 110(D24). doi:10.1029/2005jd006318
- 841
- 842 Xu, W., Wooster, M. J., Roberts, G., & Freeborn, P. (2010). New goes imager algorithms for
843 cloud and active fire detection and fire radiative power assessment across North, South and
844 Central America. *Remote Sensing of Environment*, 114(9), 1876–1895.
845 doi:10.1016/j.rse.2010.03.012
- 846
- 847 Zhou, M., Wang, J., Garcia, L. C., Chen, X., da Silva, A. M., Wang, Z., ... & Miller, S. D.
848 (2023). Enhancement of Nighttime Fire Detection and Combustion Efficiency Characterization
Using Suomi-NPP and NOAA-20 VIIRS Instruments. *IEEE transactions on geoscience and
remote sensing*, 61, 1-20.
- 849
- 850 Zou, Y., O'Neill, S. M., Larkin, N. K., Alvarado, E. C., Solomon, R., Mass, C., et al. (2019). Machine learning-based integration of high-resolution wildfire smoke simulations and

851 observations for Regional Health Impact Assessment. International Journal of Environmental
852 Research and Public Health, 16(12), 2137. doi:10.3390/ijerph16122137

PDF-PRINT