

Trial by Fire: Diachronic and Regional Sentiment Towards American Wildfires

Charles Hutchinson, Sam Kalman

December 12, 2022

Abstract

We analyzed diachronic regional trends in public tweets about wildfires to better understand public perception towards climate change. Processing over 350,000 tweets, we used both a dictionary-based approach in the `syuzhet` R package and a roBERTa model from researchers at Cardiff University for sentiment analysis. We found that sentiment has become increasingly negative over time, particularly in the Pacific Northwest. This conclusion could be reinforced by directly comparing air quality and sentiment.

Introduction

One of North America's most noticeable symptoms of climate change has been a dramatic increase in the volume and intensity of wildfires in its western regions. August and September were previously defined by clear skies and fair weather, but fire and smoke have become a recurrent factor in this area. According to the National Interagency Fire Center (NIFC), more acreage burned nationally in each of the last seven years than any single year but one in the first ten years this data was collected (1983-1993, ex. 1988).¹ With more and more people affected by the increase in wildfire activity, it's reasonable to infer an increasing outcry of negativity towards this devastating symptom of climate change. However, less is known about regional attitudes towards this recent trend. As with many other catastrophes, is it true that wildfires are "out of sight, out of mind" for those not directly affected?

We explored four recent US wildfires that took place in the past five years to analyze if and how public sentiment has shifted over time with respect to these cataclysmic events. Specifically, we sought to answer the following two questions:

1. How has public sentiment towards wildfires changed in the past five years?
2. Is this change regionalized? Are changes in sentiment (if any) limited to those directly affected by these devastating events?

We compared two different sentiment analysis algorithms in the hopes of gaining some insight into changing public perceptions of climate change.

Data

We analyzed sentiment towards wildfires by pulling publicly available tweets through Twitter's API. We chose to pull tweets made around the time of four of the largest, most recent wildfires in California. These events were the 2021 Dixie Fire, the 2020 Bay Area Fires, the 2018 Camp Fire, and the 2017 Tubbs Fire.

It was difficult to find relevant data. Tweets that were well-tagged (e.g., including #DixieFire for the Dixie Fire) tended to be associated with news organizations, which were overwhelmingly neutral and rarely "organic" reactions. However, loosening the search parameters inevitably lowered the data quality (e.g., searching for "smoke" returned a large sample of tweets referencing marijuana usage). It was especially challenging to develop searches with consistent methodology across separate wildfires, while still getting useful data.

¹See <https://www.nifc.gov/fire-information/statistics/wildfires>.

We found our best results using the following approach:

- Use the commonly accepted fire name and “fire” as search terms. For example, the search terms for the 2021 Dixie Fire would simply be “dixie fire”.
- Limit the search results to English. Creating sentiment dictionaries in other languages was out of scope for this project.
- Remove retweets.
- Limit the tweets to a month time frame. Some fires lasted longer than this, so we picked the interval in which the smoke was most impactful.

The size of each fire’s corpus was unfortunately still far from uniform, possibly due to unequal media coverage, but we found this to be a reasonable compromise among a set of unappealing options. We found both Camp Fire and Dixie Fire generated far more content on Twitter under our listed constraints when compared to Tubbs Fire and the Bay Area Fires. Table 1 shows the volume of tweets per fire.

Table 1: Tweet Counts by Wildfire.

Fire	Tweets	Words
Tubbs Fire	3530	51187
Camp Fire	50469	1021221
Bay Area Fires	4680	129082
Dixie Fire	26139	577785

Our twitter license included access to a range of metadata. We chose to analyze tweet level information (specifically Likes, Retweet, Quote Tweets, and Date), as well as certain user level information (specifically Followers, Following, Biography, and Twitter Name).

Filtering for tweets that explicitly mentioned the fire by name significantly limited the total number of results. This proved to be particularly debilitating when working on location-based analyses, since only a subset of that limited population reported their location while tweeting. To alleviate this problem, we chose to use much more generic search terms when conducting our location-based analyses. For the purposes of this paper, we found the search term “fire” tended to return the most relevant results (as compared to similarly broad terms like “smoke” or “burn”), though this was not a systematic decision. Nonetheless, it is hard to argue that “fire” is not an appropriate search term to use when conducting sentiment analysis on wildfires.

In our location-based analyses, we continued to use the same date windows for each fire. As before, we only pulled English tweets, and filtered out any retweets. We asked the Twitter API to filter for tweets in the United States, then manually filtered those tweets down to those made in the continental US with longitude and latitude filters. Twitter provides locations in a bounding box, not on a single point, so we assumed that each tweet was made in the center of the bounding box. This was somewhat problematic when dealing with state-size bounding boxes, but the vast majority of locations were given on the city level, with much smaller bounding boxes.

Table 2: Location-Enabled Tweet Counts by Wildfire.

Fire	Tweets	Words
Tubbs Fire	76839	1165880
Camp Fire	84284	1746368
Bay Area Fires	71608	1468934
Dixie Fire	36353	749509

We found that most tweets from news organization contained neutral, non-opinionated information, which wasn’t useful for our research question. As such, any tweet made by a user with the word “meteorology” in

their Twitter biography was not considered in our analysis. Additionally, in an attempt to filter out bots adding unnecessary noise to the data, we only used tweets from users with at least one follower, and the tweet itself needed at least one external interaction (a like/retweet/comment). It's worth noting that the authors of Cardiff's roBERTa model (Loureniro et al., 2022) addressed the bot issue by filtering out the top 1% of users by volume, but we did not have a similarly unbiased data set to determine an appropriate posting cutoff.

Methods

We used two sentiment analysis algorithms of varying complexity to classify the aforementioned tweets. The first algorithm we used was from the `syuzhet` package² in R. The `syuzhet` package returns sentiment scores on a spectrum of positively/negativity using a sentiment dictionary developed in the Nebraska Literacy Labs. We deemed any sentiment score that was greater than zero to be positive, less than zero, negative, and exactly equal to zero, neutral. Unfortunately, the `syuzhet` package wasn't designed for or trained on short, declarative statements such as tweets.

To overcome this issue, we decided to use Cardiff's roBERTa-base model for sentiment analysis as well.³ This model was trained on roughly 124 million tweets from January 2018 to December 2021, using the TweetEval benchmark to fine tune the sentiment analysis. The model outputs probabilities that a tweet is positive, neutral, or negative. For the purposes of our analysis, we simplified the output to a categorical variable representing the sentiment with the highest probability. (Strength of sentiment was not captured.)

BERT models are trained on both language modeling and next sentence prediction. It is important to note that BERT models generally do not use words as tokens - they use a combination of word-level, segment-level, and position-level tokens to encode the input text. Word-level tokens are individual words or sub-words in the input text, segment-level tokens are used to indicate whether a given word is part of the first or second sentence in a pair of sentences (in the case of sentence-pair tasks), and position-level tokens are used to indicate the position of each word within the input text. This may convey some of the complexity of the contextual embedding performed in the model. Regardless, we note that the word counts in Tables 1 and 2 are not accurate token counts for Cardiff's roBERTa model.

To track to sentiment shift through time, we chronologically ordered the fires (Tubbs Fire, Camp Fire, Bay Area Fires, Dixie Fire) and calculated the proportion of each sentiment for each fire. We used proportions because of the unequal distribution of tweets across the fires. We apply the Bonferroni correction to our 95% confidence intervals to account for multiple testing.

Plotting sentiment by location proved to be a more involved task. As stated in the Data section, we chose to simplify the provided bounding boxes for each tweet to the center point. From there, we converted from longitude/latitude coordinates to WGS84 coordinates, then projected these coordinates using a Lambert Conformal Conic projection (LCC). This process was completed using the `sf` package. Using these transformed coordinates and the corresponding sentiments, we used weighted k-nearest neighbors with a Gaussian kernel to interpolate probabilities of a given sentiment across the map.

Results

Change Over Time

In Figure 1, we plot the sentiment classification proportions for each of the four wildfires. The fires are in chronological order, beginning with the Tubbs Fire and ending with the Dixie Fire. The error bars reflect a 95% confidence interval for each proportion. There are not many neutral tweets, though our definition of neutrality with the `syuzhet` scoring is quite stringent. The majority of tweets are classified as negative, and the proportion of negativity increases over time.

We repeat the same analysis using scores from Cardiff's roBERTa model in Figure 2. The roBERTa model classifies a much larger proportion of the tweets as neutral. In previous iterations of this analysis, it was difficult for us to remove all news related tweets. We found that this iteration was very effective at putting

²<https://www.rdocumentation.org/packages/syuzhet/versions/1.0.6>

³<https://huggingface.co/cardiffnlp/twitter-roBERTa-sentiment-latest>

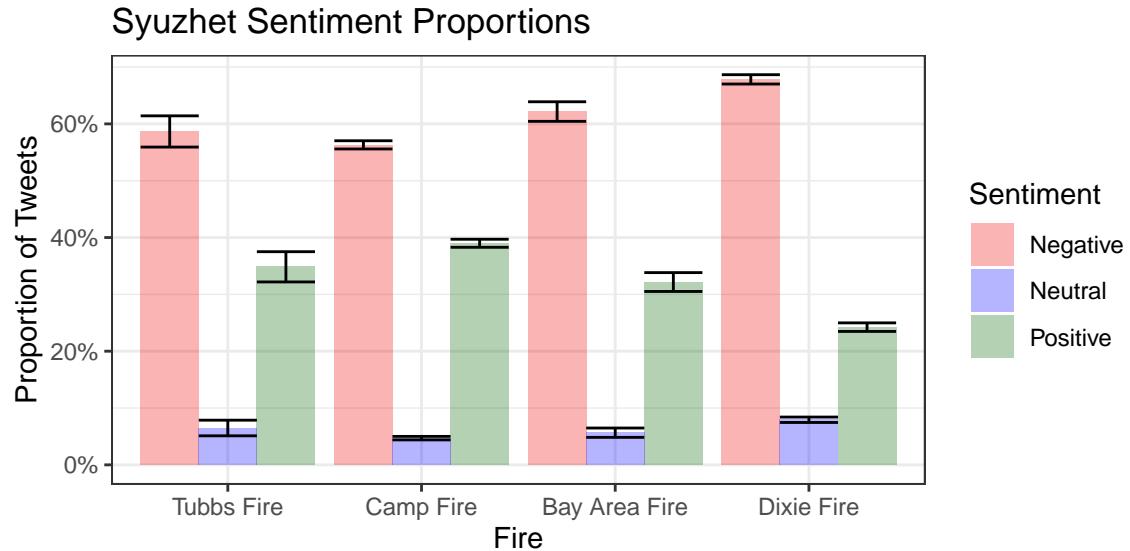


Figure 1: Shift in Wildfire Sentiment Over Time.

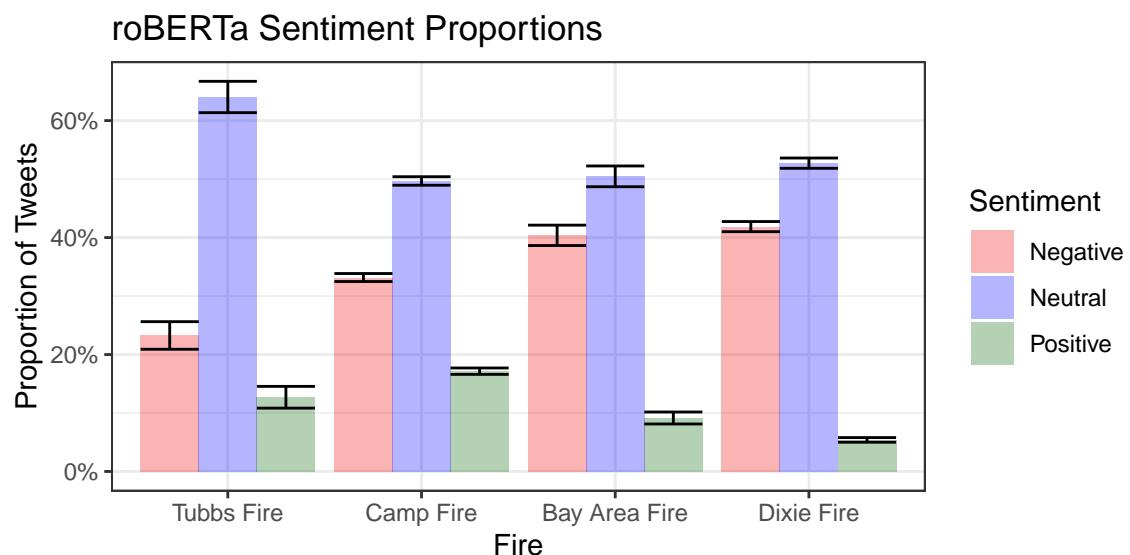


Figure 2: Shift in Wildfire Sentiment Over Time.

“inorganic” reactions from news organizations in this neutral category. Additionally, the roBERTa model classified less negativity for the Tubbs fire than the *syuzhet* package and we see a more consistent increase in negativity, plateauing at the Bay Area Fires and Dixie Fire.

Effect of Location

Figure 3 shows the interpolated roBERTa sentiment probabilities across the continental US. The obvious trend is one of consolidation - in particular, the Pacific Northwest has become increasingly negative about fires during fire season. As one of the hardest hit regions by these catastrophes in the country, this is not a surprising result.

It is important to reiterate that these maps use a more generic data set than the bar plots above. Specific mention of the fire’s name was not required - only the term “fire” was needed. There are a wide variety of ways that the term “fire” is used on Twitter, many of which are not related to wildfires. Manual inspection revealed that this accounts for the large swaths of positive tweets in plots such as Figures 3a and 3d. However, we were unable to find a robust explanation for the regional variation in positivity over time.

The intensity of color in Figure 3 corresponds to probability of a given sentiment, not volume. Plotting by the latter would result in a mostly blank map, with a few spots of color around urban centers.

Comparison of Models

While performing manual checks of each model’s sentiment classification, we were consistently impressed by the performance of Cardiff’s model. The BERT framework was able to capture context in a way that was impossible for *syuzhet*’s dictionary-based approach. We elaborate on some common cases we saw below.

1. Neutral roBERTa + Negative *syuzhet*

Where Did the Tubbs Fire Begin? Investigators Scour the Ashes <https://t.co/9z2QjBXi77>

The cause of the massive Tubbs fire is still unknown – but officials think they may be zeroing in: <https://t.co/0k9w4nFgpE>.

These are examples of neutral tweets from news organizations. These were incorrectly classified as negative by *syuzhet*, due to words like “scour” or “unknown”, though we note that the requirement of a zero sentiment score for neutrality contributed to this incorrect classification.

2. Positive roBERTa + Negative *syuzhet*

This Friday’s midnight showing the Rocky Horror Picture Show is now a benefit for Tubbs Fire victims: <https://t.co/Su7tOzcnfm>

All I can say is wow. Props to Berkeley #firefighters. Crazy video of firefighters stunned by scope of Tubbs Fire <https://t.co/8kZwh6g5uy>

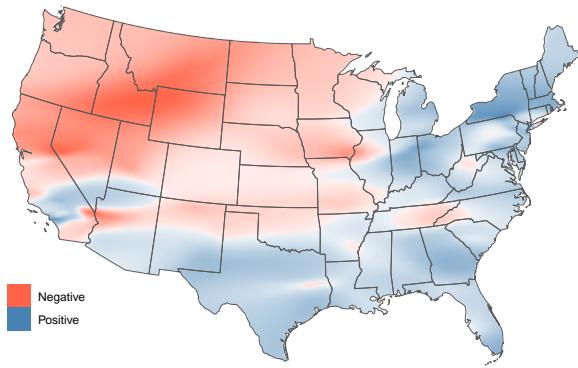
In the first tweet, *syuzhet* is misled by the title of the movie (Rocky Horror Picture Show), while roBERTa identifies the altruistic context (a charity showing). In the second tweet, *syuzhet* overweights the negative language with respect to the fire, while roBERTa correctly captures the heroic context of the tweet. These examples highlight how the roBERTa model was more effective at distinguishing subtle nuances in tweets, resulting in more accurate classifications of sentiment.

3. Negative roBERTa + Positive *syuzhet*

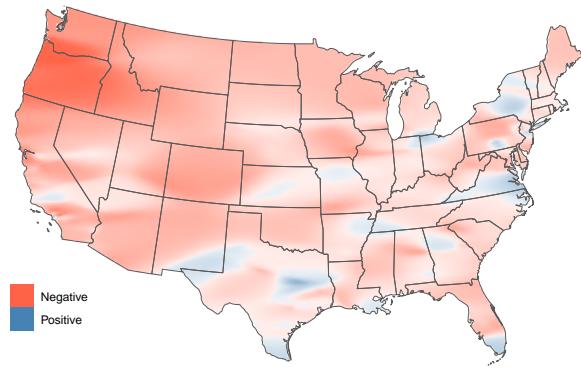
The personal treasures former Raiders great Cliff Branch lost in Santa Rosa fires—a truly sad tale by Bob Padecky. <https://t.co/gIT0AulURq>

My parents lost everything in the Tubbs Fire yesterday. Anything will help. Please share as much as you can. Thank you so much. #tubbsfire <https://t.co/xFAexYshSL>

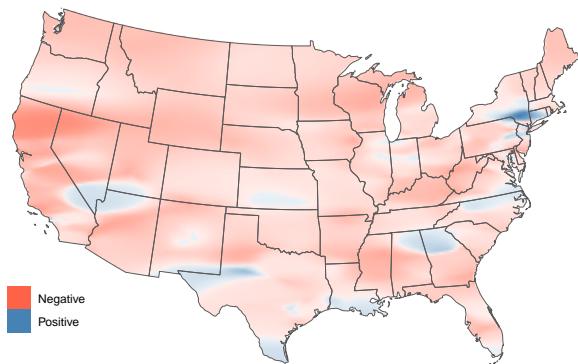
Many Sonic staff in Santa Rosa and surrounding area are affected by the Tubbs Fire. Best wishes to all, and be safe! It is devastating. <https://t.co/pVqAFisxKd>



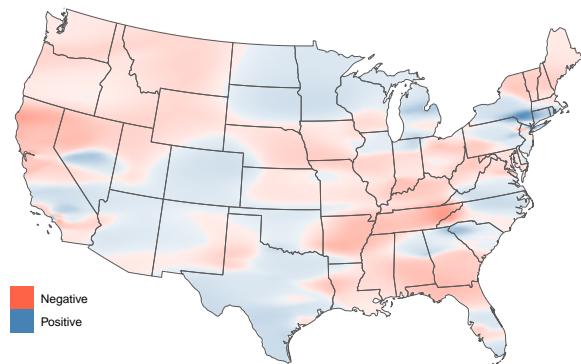
(a) Dixie (2021)



(b) Bay Area (2020)



(c) Camp (2018)



(d) Tubbs (2017)

Figure 3: Tweet Sentiment (roBERTa) by Location.

The first tweet uses the word “great” to describe the football player Cliff Branch, but the overall sentiment should be negative due to Branch losing his possessions in the fire. Many tweets containing the terms “thank you” or “be safe” were seen as positive by the `syuzhet` model, when the overall message of the tweet was in a negative context. For the purposes of our analysis, this distinction was the most important improvement between roBERTa and `syuzhet`’s output.

Discussion

Figures 1 and 2 shows evidence of a negative shift in sentiment towards wildfires in the last five years. The increase in negativity is broadly consistent between sentiment models. Figure 3 shows an increasing concentration of negative sentiment towards fire in the Pacific Northwest. We propose that negative sentiment towards wildfires has become increasingly localized over the past five years.

As mentioned previously, data quality was the limiting factor for this analysis. It was difficult to find a set of relevant tweets using a consistent methodology for each fire. Finding a balance between organic and relevant reactions ended up requiring more judgment than we would have liked.

In our first data set (Table 1), sample size varied by more than an order of magnitude between certain fires. It was surprising that we could not find more tweets for the Bay Area and Tubbs fires, given their proximity to San Francisco. We originally suspected that the Camp Fire tweet set contained many false positives, but it appears that fire generated more engagement than expected due to the almost total devastation (and subsequent press coverage) of the towns of Paradise, Concow, Magalia, and Butte Creek Canyon. The Dixie Fire was similarly destructive. As the largest ever single wildfire in the state’s history, part of its impact included popular Lassen National Park, the coverage of which likely increased engagement on Twitter.

Given more time, it would have been helpful to correct for fire intensity. These fires were all record-setting, but they took place in different conditions. Some were closer to urban centers, and some were quite far. Some took place in the middle of summer and were coincident with a large set of other wildfires, while others (particularly the Camp Fire) took place later in the season, in near isolation. Early September of 2020, near the peak of the Bay Area fires, was a particularly striking example of how catastrophic these coincident fires could be (see Figure 4). It would be particularly interesting to compare sentiment to air quality, though we suspect the combination of (typically) rapid fluctuations in air quality and the sparsity of location-enabled tweets on any given day would make robust statistical conclusions difficult.

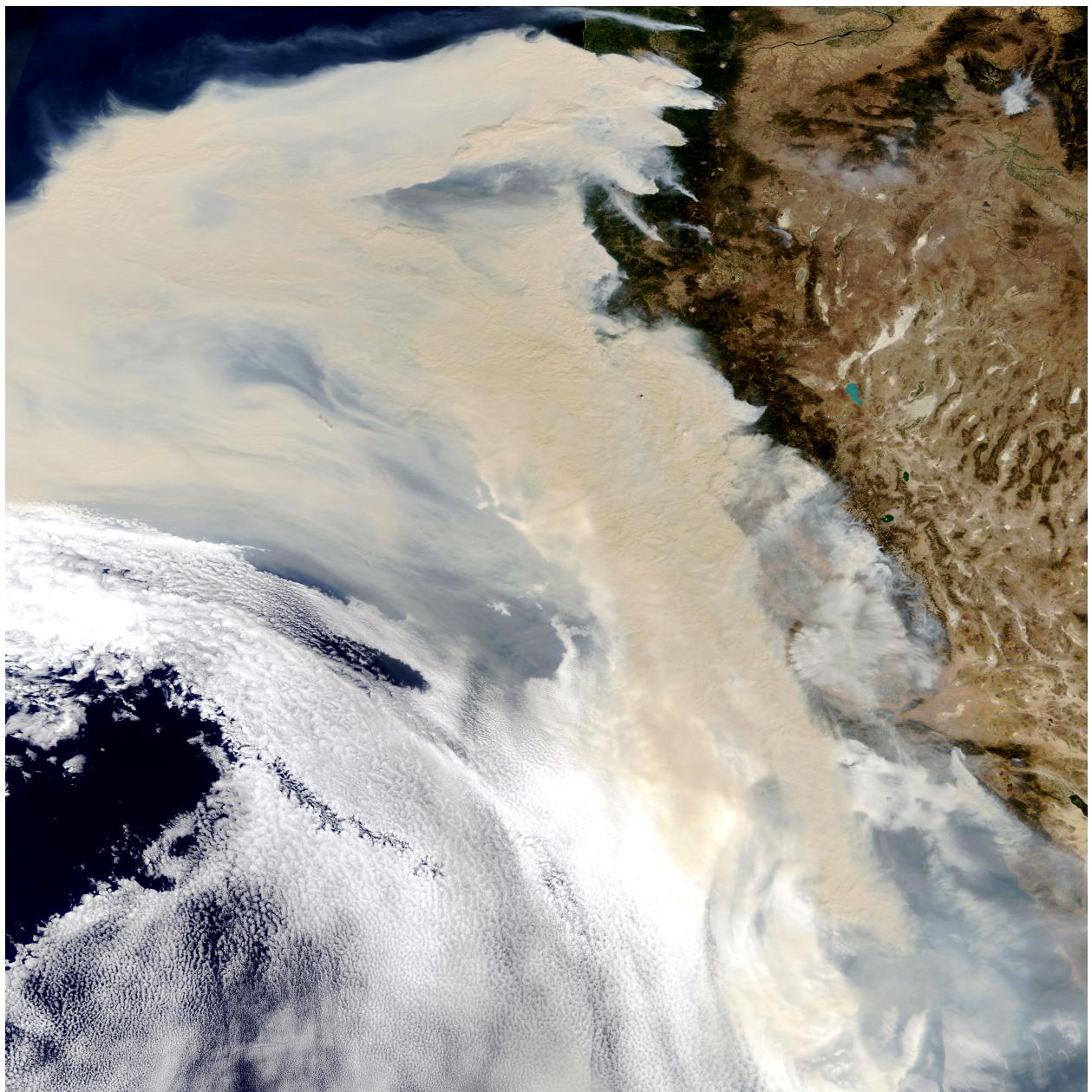


Figure 4: Wildfire Smoke as of September 9, 2020. By NASA Earth Observatory.