Response to reviewer

This paper uses long-term Aqua and Terra MODIS Fire Radiative Power (FRP), Aerosol Optical Depth (AOD), and surface observations of PM2.5 (particulate matter with median diameter smaller than 2.5 um) to study the impact of smoke from wildfires on surface PM2.5. The authors picked a 2-week time period in 2018 to represent extreme wildfires and 2011 to represent low wildfire activity to compare and contrast and report the impact. While it is true that the human induced pollution levels are going down in the US and the role of natural events such as wildfires and dust storms in influencing the air quality is increasing, I find this work very rudimentary and without scientific rigor. The paper is well written no doubt but this work is merely an exercise of downloading data from different sources and making figures. Let me explain why I think this study needs a major re-work (scientific scope as well as methodology) and is not ready for publication.

We thank the reviewer for the extensive and detailed review of our manuscript. We believe the comments improved the paper, and we have revised the paper significantly in light of the reviewer's suggestions. Our point-by-point response to the reviewer's comments is given below:

First and foremost, to conduct this study, there is no need to use satellite data because the analysis is done in an aggregate sense, spatially and temporally. There are enough ground monitors in different states influenced by smoke from fires that a study can be designed just around the surface monitors without even bringing in the errors associated with scaling satellite AOD to surface PM2.5;

While there would be less errors without using satellite AOD, pollution cannot be estimated at places that lack ground monitors such as Wyoming (shown in figure 1 below). We also added some explanation in the discussion section.

Canadian wildfires in some years affected the US east coast, and the smoke can be captured by ground monitors. However, wildfires in summer 2018 mostly influenced the northern and western part of US where, unlike eastern coast, population is less. Ground monitors of many affected states are only concentrated in the few population centers and leave large gaps of PM2.5 observations in these states. For example, there are 13 EPA PM2.5 stations in Wyoming state (figure 1a), but they are distributed in two corners, while leaving a large portion of the state unmonitored. It is similar a situation in other states except for Washington and California State.

It is possible to estimate PM2.5 by applying interpolation methods on the ground observations but not feasible when there are large gaps. Also we decided to use GWR method and utilize different variables to increase the prediction accuracy.

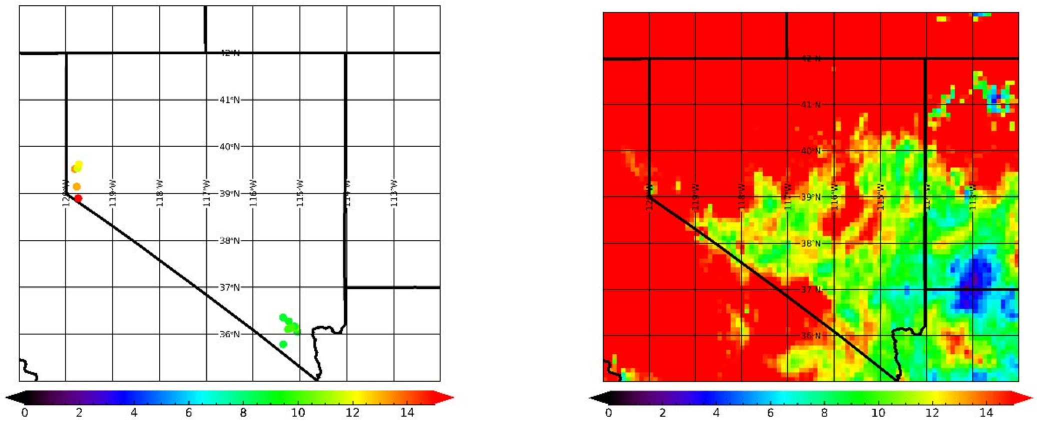


Figure 1. 17-day mean PM2.5 distribution from (a) EPA ground monitors (b) GWR estimation

Second, some of the stations (100s of them) have daily (or every other day) speciation measurements including organic carbon and K+, biomarkers for smoke. The authors have not bothered to analyze the surface data and extract only data for the days or locations influenced by smoke. Yes, there are speciation observations from EPA network as well as interagency network (IMPROVE) in many of the states where smoke originates and many states downwind of smoke;

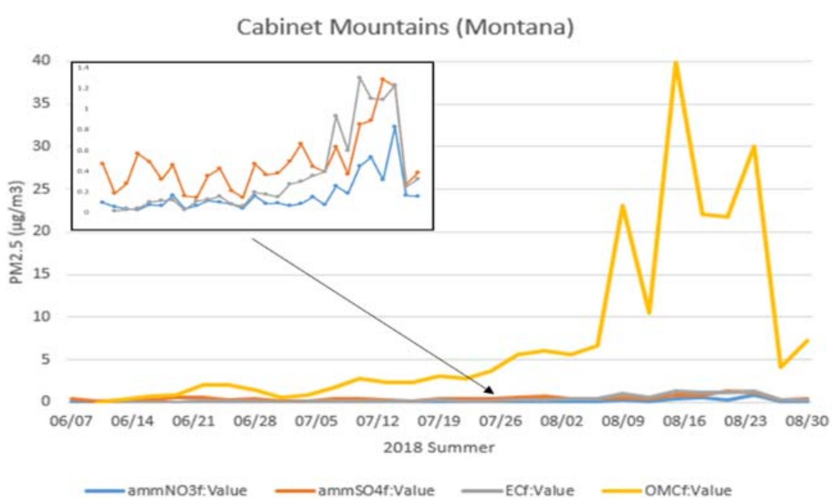


Figure 2. IMPROVE data.

We thank the reviewer for pointing this out. We actually checked selected ground observations, and assessed that high PM2.5 indeed originated from fires. We did not intend to filter the data to analyze only for smoke pixels since an estimation of smoke affected regions also needs unaffected pixels to feed the model.

We checked a few IMPROVE sites of the affected states, and found high organic carbon mass concentration and noticeable increases in elemental carbon, ammonium nitrate and ammonium sulfate. Figure 2 is an example of one site in Montana State (115.6709°W, 47.9549°N). We also compared the EPA PM2.5 distributions with other species from the EPA stations (Figure 3), and we conclude that the increase of PM2.5 is due to wildfires since high values of PM2.5, elemental carbon PM2.5 and organic carbon PM2.5 all distributed the same.

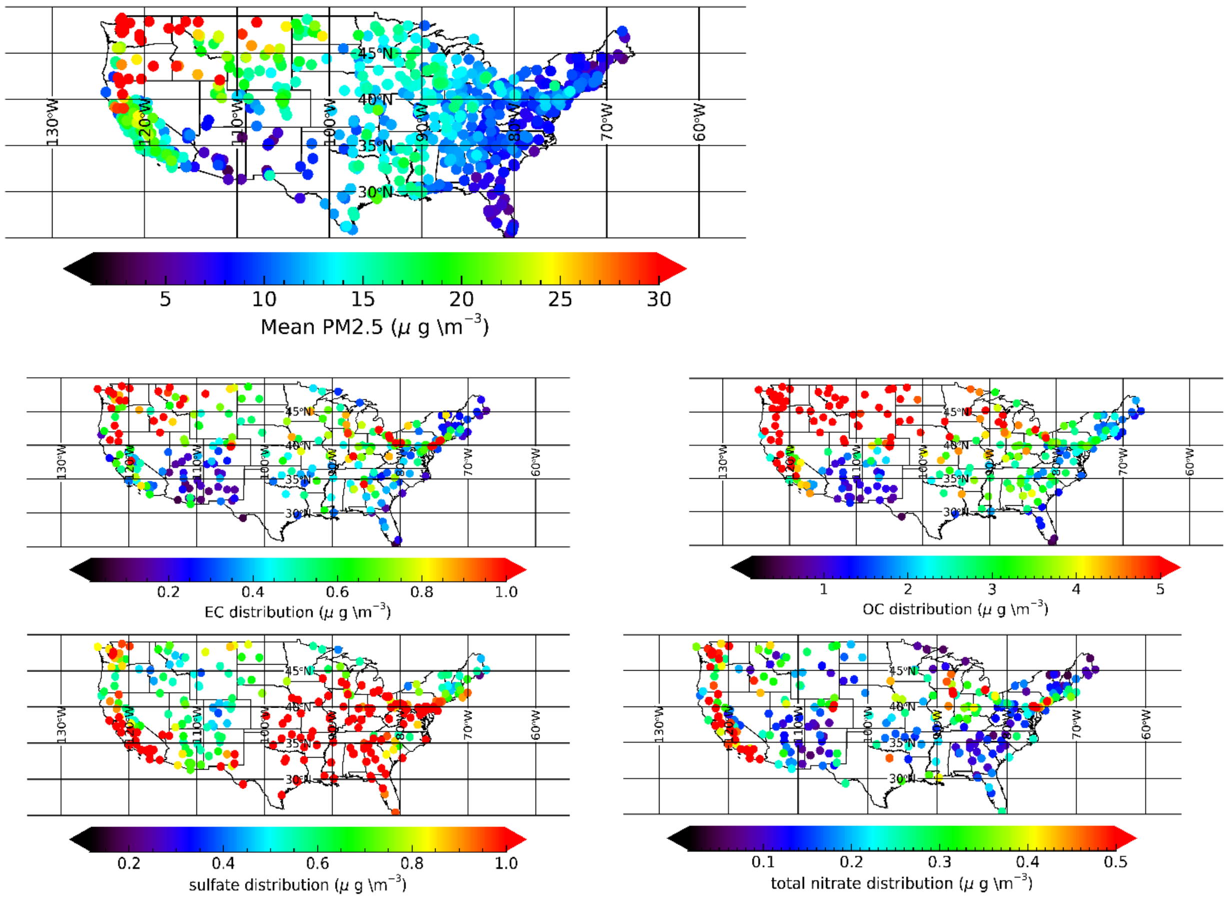


Figure 3. Surface Distribution of chemical speciation

Third, if one or more ground monitors in a county/state are influenced by upwind smoke from fires, an exceptional events waiver must have been filed with the EPA. Did the authors check to see how many exceptional events waivers were filed for 2018 by the states that were under the smoke influence as reported by the authors?

We checked the exceptional events waivers for 2018 (Figure 4). The results are consistent with our findings: the affected states from our analysis had indeed more exceptional events waivers.

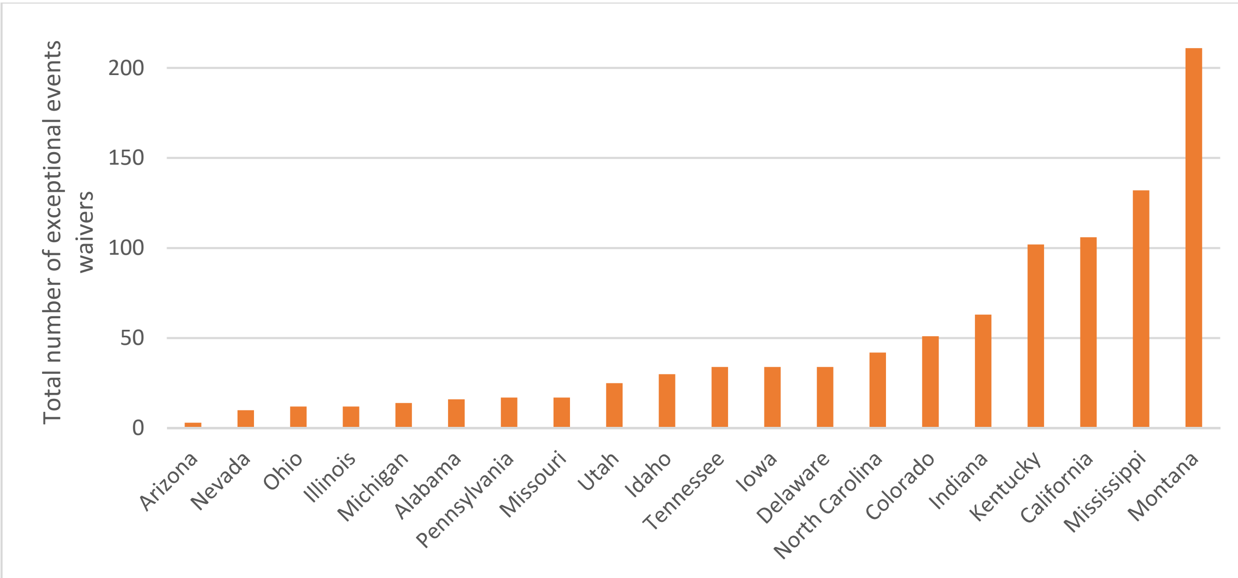


Figure 4. Exceptional Events

Figure 4 shows the number of exceptional events waivers filed during the study period (August 9th to August 25th, 2018) in different states. States with no exceptional events waivers during this period are not shown in the figure. The values are not only related to the number of EPA sites in each states but also relevant to the distribution (location) of these sites.

Given that there are fire observations (ground reports from EPA as well as from satellites) and surface PM2.5 data for two decades, why not conduct or extend the study to all years to understand the nuances of the inter-annual variability and the influence of transport etc. Again, this is why I find this paper very premature because the authors have not even scratched the surface of the problem.

It would have been interesting to explore this aspect. However, the purpose of this paper is to apply GWR method on a selected severe wildfire event, test its predicting performance over region that contains high concentration smoke, and then quantify the influence of one wildfire event on the US air quality. Studying the variability of past 20 years' wildfires is beyond the scope of this paper.

There are many documented algorithms that use satellite data to flag smoke and smoke height including the MAIAC aerosol algorithm used in this study. The authors used AOD but not smoke flag and smoke plume height product generated by the same algorithm. While the smoke plume height product is new, the smoke flag and AOD in the MAIAC algorithm are internally consistent and the authors should have used it in this study. Also there is no discussion on the quality of the MAIAC AOD and its performance. The algorithm performance is reported as 66% of the retrievals are within 0.5? I am not exactly sure why this is a good performance? How good is the AOD product in different AOD ranges? Does it report AODs as high as 5 or 7 for these smoke events or smoke is misidentified as cloud? If an aerosol model is used in the algorithm, does the algorithm dynamically (correctly) pick smoke model for this time period? How consistently does it pick the smoke model? If another model is picked, what is the AOD bias for incorrectly picking a non-smoke model? And how does that translate to PM2.5 estimation error?

The smoke flag is added to the GWR model as a predictor and we have revised the reference for MAIAC AOD performance. Over North America, MAIAC AOD has a very small bias of -0.01 compared to AERONET AOD (Superczynski et al., 2017). The typical error is usually around ±0.05 during times of high aerosol loading, and the bias slightly increases as AOD increases.

The MAIAC AOD product has a maximum value of 4, which should be enough for our study since the major fire sources in Canada are far away from the US. The smoke detection is performed using MODIS red, blue and deep blue bands, and separate smoke pixels from dust and clouds based on absorption parameter, size parameter and thermal threshold. For now, there is no explicit biomass burning aerosol models included in the MAIAC retrievals. For pixels with no smoke detected, upper 50% of the data will be filtered as potentially affected by clouds or shadows, which will possibly lead to missing data.

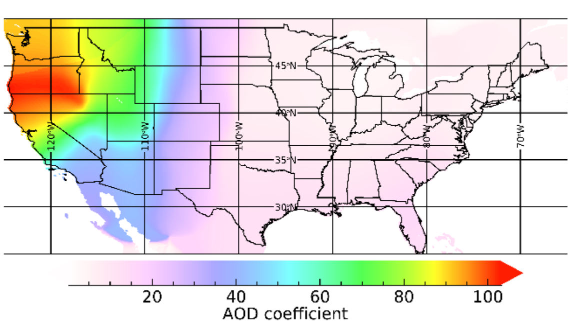
Please show a map of regression parameters and demonstrate that the values have physical meaning

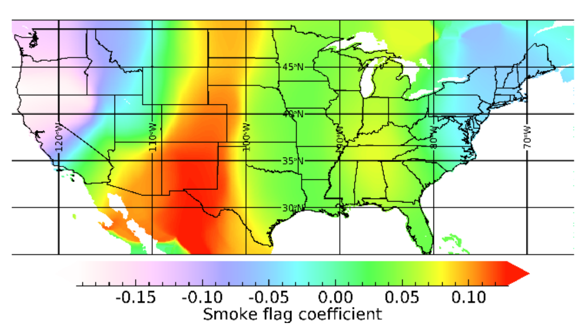
The figures below show the distribution of some regression coefficients. AOD coefficients are greater close to the fire sources (Northwestern US) and gradually decreases with distances increase, which means AOD is more dominant in predicting PM2.5 near the fire.

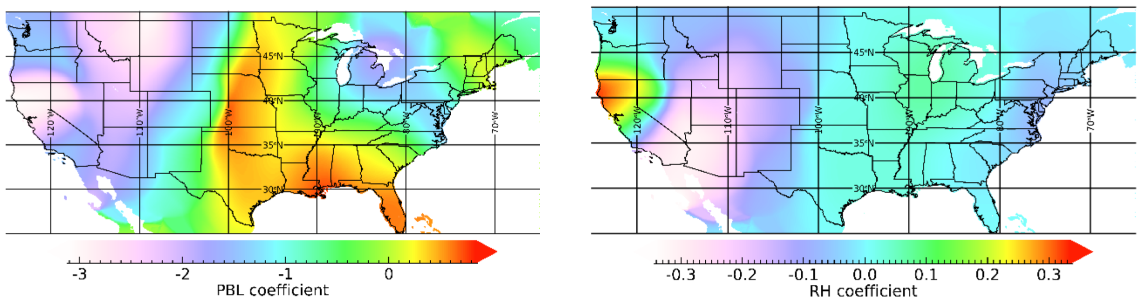
The smoke flag is overall positive related to surface PM2.5, while it could slightly negatively relate to PM2.5 around fire sources and northeastern coasts.

The PBL are negatively related to PM2.5 when the pollution is concentrated around the surface (fires or human-made emissions), while it appears to be positive related to PM2.5 at locations where the main pollution source comes from remote wildfire smoke.

Relative humidity, on the other hand, shows large variations on PM2.5 influence across the nation. Around the wildfires where the RH is relative low, RH has a positive correlation with PM2.5 since hygroscopic would increase and leads to accumulation of PM2.5, but increasing RH can also decrease PM2.5 concentration by overgrowing the PM2.5 particles to deposition at high RH environment.







give details on why you chose the parameters you chose for the model. Let us talk about population density. Why did you use it? I can understand why it is used if you are developing models for urban/industrial pollution where population density can be a proxy for traffic emissions etc. Here, isn't the focus of the study to understand the influence of long-range transport of smoke from fires on humans and their health. Then how can population density be a predictor?

The reviewer is correct, and we have removed population density from the GWR model. More details on predictor choosing is described in Data section.

no details given on the influence of different predictors such as boundary layer height on the prediction

We have added some explanation on how predictors can influence the PM2.5 based on the coefficient's distribution (section 4.3).

The authors have not shown their assessments on how good the estimated PM2.5 values are outside of one evaluation (scatter plot for the whole US). If you look at the density of the data points, most points are within 0 to 20 ug/m3 or so. When the EPA PM2.5 daily average standard is 35 g/m3, I would be more interested in knowing the performance of the statistical model for exceedances. Can the authors actually tabulate what percentage of each jurisdiction (e.g., state) violated the daily standard and how many times within the 2-week window in 2018?

We think this is an excellent suggestion. For data greater than 35 $\mu g\ m^{-3}$, the model has a RMSE of 12.07 $\mu g\ m^{-3}$, which is a lot larger than the whole model RMSE. Therefore, the model has a tendency for underestimating PM2.5 exceedances by around 12.07 $\mu g\ m^{-3}$. The larger the PM2.5 is, the greater the model underestimates.

Also, in our study, it would not be possible to calculate the number of days that violating the EPA standard, because we estimate surface PM2.5 over a 17-day period, not daily estimation. But we add some analysis (with below table) using ground observations in the discussion section.

state	number of site violate standard	number of site in the state	Percentage of site violate standard (%)	number of days violate standard
Montana	14	15	93.34	16
Washington	18	20	90	16

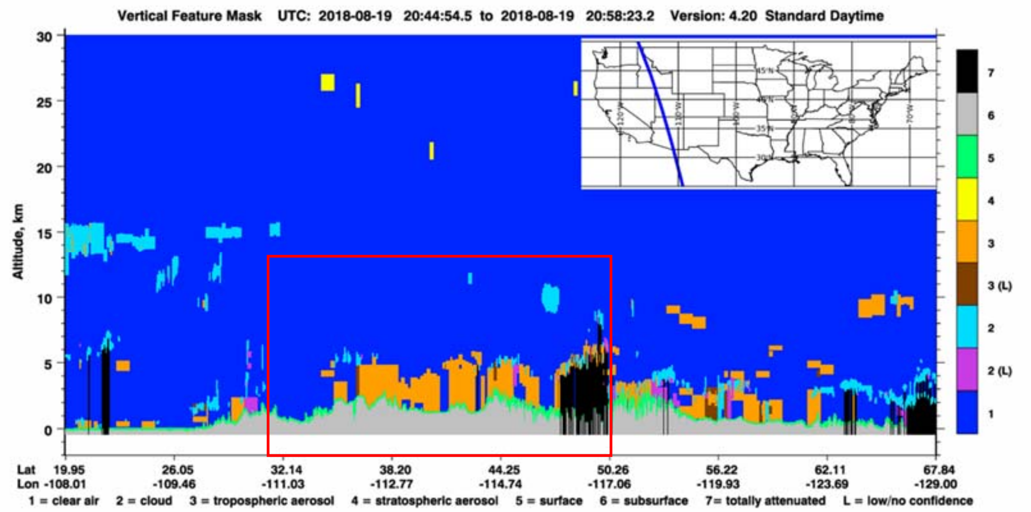
Oregon	12	14	85.71	5
North Dakota	7	11	63.63	4
Idaho	5	8	62.5	8
Colorado	11	21	52.38	2
South Dakota	5	10	50	1
California	57	119	47.9	14
Utah	7	15	46.67	4
Nevada	4	13	30.77	1
Wyoming	7	24	29.2	2
Minnesota	4	26	15.4	2
Texas	3	37	8.1	1
Louisiana	1	14	7.1	1
Arizona	1	20	5	1

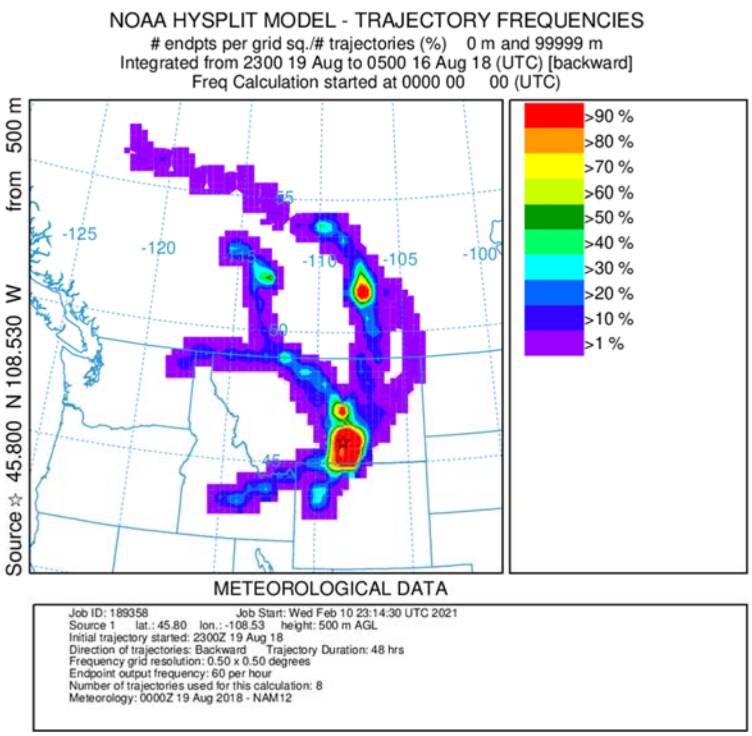
this study also needs other corroborative evidence such as back trajectory cluster analysis to show the source-receptor relationship, analysis of LIDAR data (satellite or ground) to show evidence of transported smoke reaching the surface etc.

Based on the reviewer's suggestion, and we checked for several different datasets for the existence of smoke reaching surface and back-trajectory paths to find the smoke is indeed originated from both local and remote fire sources.

Below Figure shows the vertical feature mask from CALIPSO on August 19th 2018, and the blue line in the inset map on the top right corner represent for the satellite orbit track. Within US (shown in the red square box), there are large load of aerosols below 5km along the track (Idaho, Utah and Arizona).

The second figure shows the pollution (smoke) at Billings (Montana) on August 19th was originated from both remote fires in Canada and local fires in Washington and Idaho.


Vertical Feature Mask UTC: 2018-08-19 20:44:54.5 to 2018-08-19 20:58:23.2 Version: 4.20 Standard Daytime
Altitude, km
30
25
20
15
10
5
0
Lat 19.95 26.05 32.14 38.20 44.25 50.26 56.22 62.11 67.84
Lon -108.01 -109.46 -111.03 -112.77 -114.74 -117.06 -119.93 -123.69 -129.00
1 = clear air 2 = cloud 3 = tropospheric aerosol 4 = stratospheric aerosol 5 = surface 6 = subsurface 7= totally attenuated L = low/no confidence
7
6
5
4
3
3 (L)
2
2 (L)
1


NOAA HYSPLIT MODEL - TRAJECTORY FREQUENCIES
endpts per grid sq./# trajectories (%) 0 m and 99999 m
Integrated from 2300 19 Aug to 0500 16 Aug 18 (UTC) [backward]
Freq Calculation started at 0000 00 00 (UTC)
Source ☆ 45.800 N 108.530 W from 500 m
-125
-120
-115
-110
-105
-100
55
50
45
>90 %
>80 %
>70 %
>60 %
>50 %
>40 %
>30 %
>20 %
>10 %
>1 %
METEOROLOGICAL DATA
Job ID: 189358 Job Start: Wed Feb 10 23:14:30 UTC 2021
Source 1 lat.: 45.80 lon.: -108.53 height: 500 m AGL
Initial trajectory started: 2300Z 19 Aug 18
Direction of trajectories: Backward Trajectory Duration: 48 hrs
Frequency grid resolution: 0.50 x 0.50 degrees
Endpoint output frequency: 60 per hour
Number of trajectories used for this calculation: 8
Meteorology: 0000Z 19 Aug 2018 - NAM12