Estimating Wildfire Smoke Concentrations During the October 2017 California Fires Through BME Space/Time Data Fusion of Observed, Modeled, and Satellite-Derived PM2.5: Supporting Information

*Stephanie E. Cleland, J. Jason West, Yiqin Jia, Stephen Reid, Sean Raffuse, Susan O’Neill, Marc L. Serre*

This supporting information contains 24 pages, 4 tables, and 14 figures.

**Table of Contents**

[S.1 Location of PM2.5 Monitoring Stations 3](#_Toc49769352)

[S.2 AOD to PM2.5 Conversion 3](#_Toc49769353)

[S.3 CAMP Correction Figures 4](#_Toc49769354)

[S.4 Additional Details on BME 6](#_Toc49769355)

[S.5 Covariance of PM2.5 Data 10](#_Toc49769356)

[S.6 Additional Details on Global Offsets 12](#_Toc49769357)

[S.7 Equations Used for Performance Evaluation 14](#_Toc49769358)

[S.8 State-Wide All-Month Performance Statistics 15](#_Toc49769359)

[S.9 Refinement of Plume Shape with Addition of Temporary Data 17](#_Toc49769360)

[S.10 Stratified Performance Statistics for Addition of Temporary Data 18](#_Toc49769361)

[S.11 Further Comparison of BME Methods 19](#_Toc49769362)

[S.12 Animation of Estimated PM2.5 Concentrations 21](#_Toc49769363)

[S.13 Variance Maps of Estimated PM2.5 Concentrations 21](#_Toc49769364)

[References 22](#_Toc49769365)

## S.1 Location of PM2.5 Monitoring Stations

Figure S1 shows the locations of the 163 FRM/FEM and temporary monitoring stations across California that we used in our analysis of the October 2017 wildfires.

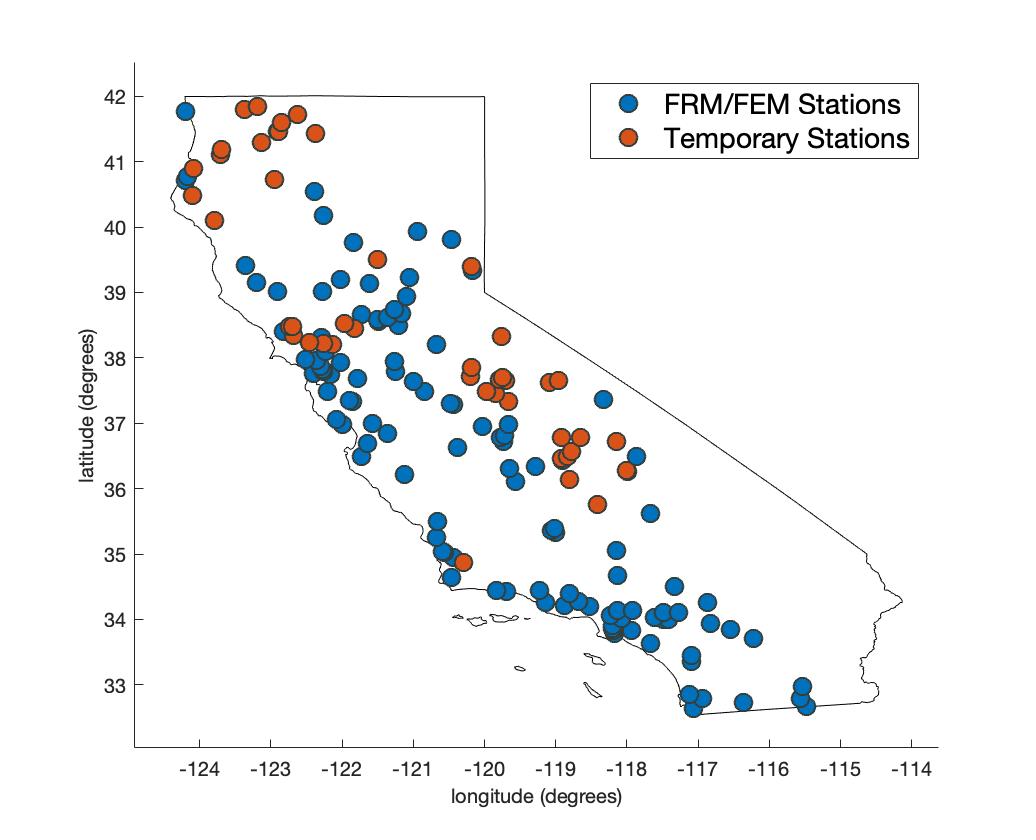


Figure S1. Locations of the 114 EPA FRM/FEM and 49 USFS temporary monitoring stations used in the study.

## S.2 AOD to PM2.5 Conversion

As detailed in the Methods section, we used a day-specific linear mixed effects model (MEM) to convert the MODIS AOD data to PM2.5 concentrations. To ensure there were a sufficient number of collocated AOD and PM2.5 observations for each day to fit the MEM, we used data across all of California. Figure S2 shows the matched, collocated AOD and PM2.5 observations for October 8, 10, and 12 and the day-specific linear MEM that was fit to the data. This figure emphasizes that the AOD-PM2.5 relationship varied daily during the fires. The equation for the MEM is:

|  |  |  |
| --- | --- | --- |
|  |  | (S1) |

where the fixed intercept and the fixed slope . There is a different value for and , the random intercept and slope, for each day. When used to convert all AOD values to PM2.5 estimations, the final MEM results in an R2 of 0.323 in the fire-affected region during the fire period and an R2 of 0.367 across all of California, October 1-31.

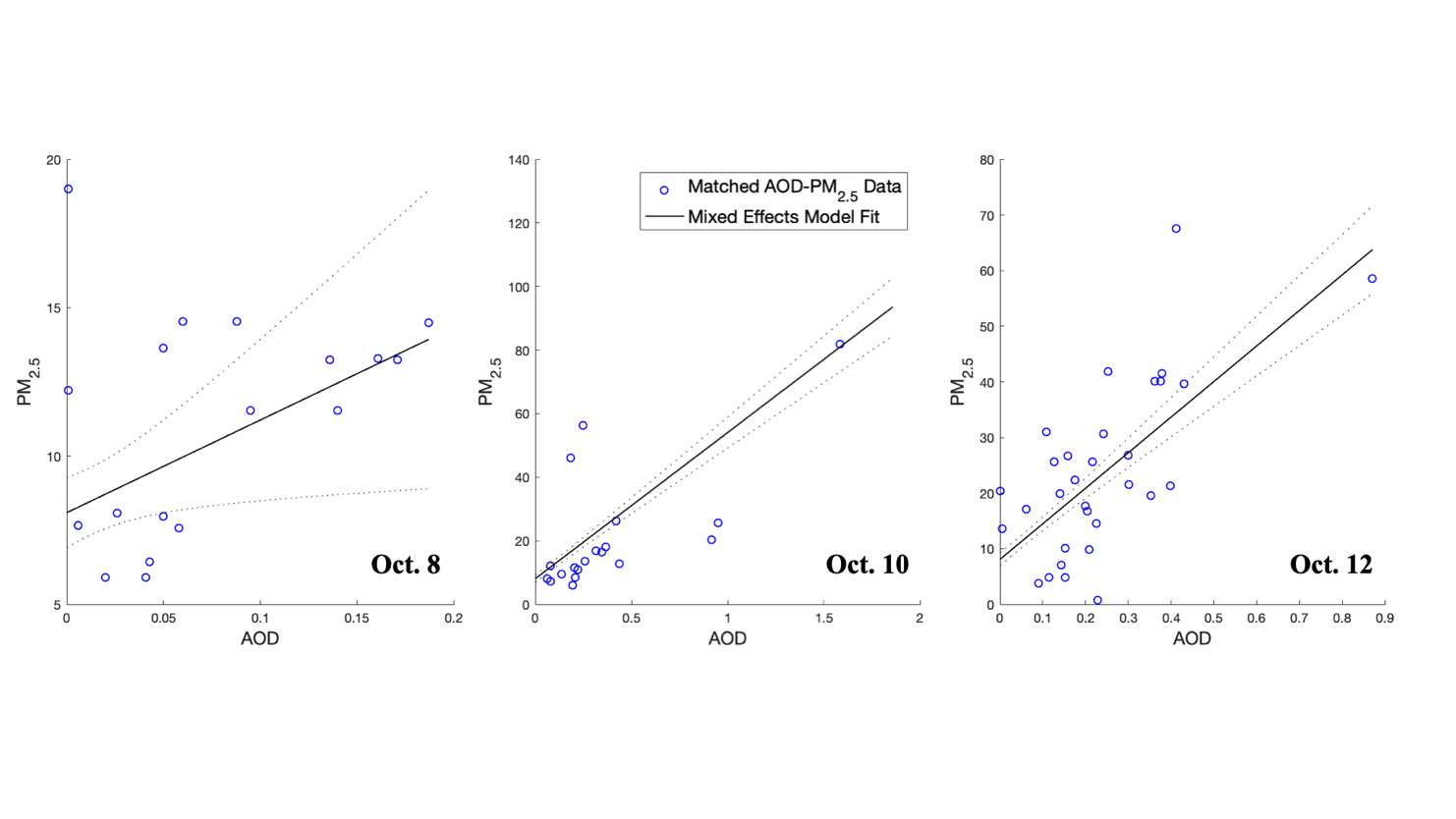


Figure S2. Day-specific linear mixed effects model, shown for Oct. 8, 10, and 12, fitted to collocated AOD and daily average PM2.5 observations.

## S.3 CAMP Correction Figures

Figure S3 shows the two CAMP correction piecewise linear functions used to bias-correct the CMAQ modeled log-PM2.5 concentrations in (1) the fire-affected region during the fire period and (2) in all other regions on all other days. The ten decile bins are shown, along with the mean ( and variance ) for each bin. The one-to-one line emphasizes the non-linear relationship between observed and modeled log-PM2.5 concentrations. For a given CMAQ modeled log-PM2.5 value, on the x-axis, the CAMP-corrected value is determined by the piecewise linear curve’s value on the y-axis. The differences between the two piecewise linear functions emphasize how the CMAQ model biases varied inside versus outside the fire-affected region and period. Figure S4 shows the impact of CAMP-correcting the modeled PM2.5 concentrations on October 10. Once CAMP-corrected, there is a notable change in the CMAQ concentrations, where extreme high values are dampened. The R2 for the CAMP-corrected CMAQ model is 0.496 in both the fire-affected region and period and across all of California, October 1-31.

#### 

Figure S3. CAMP correction piecewise linear functions used to bias-correct CMAQ modeled log-PM2.5 concentrations in (1) the fire-affected region during the fire period and (2) in all other regions and days.

#### 

Figure S4. Example of CAMP correction on Oct. 10, 2017. CMAQ daily average PM2.5 concentrations (1) prior to CAMP correction and (2) after CAMP correction.

## S.4 Additional Details on BME

The mathematical implementation of BME, adapted from Xu et al1:

There are two distinct aspects to the framework: One is a deterministic transformation of air pollution into a transformed variable, and the other is to apply the BME estimation method to estimate the value of that transformed variable across space and time.

For the deterministic transformation we use an offset-removed log-transformation, as commonly done in recent studies (see Xu et al1 and references therein). The notation used here consists in denoting a single random variable using a capital letter, e.g. , its realizations using a lower case, e.g. , and denoting vectors in bold face (e.g., ). We use the letter for PM2.5, for log-PM2.5, for offset-removed log-PM2.5. To define the transformation, we start by defining the vector of air pollution concentration observations and the vector of air pollution log-concentration observations, where the subscript denotes the case where observations are accurate, hence treated as ard data. We then define the offset-removed log-transformed data as

|  |  |  |
| --- | --- | --- |
|  |  | (S2) |

where is any deterministic offset that captures the trends in and can be mathematically calculated as a function of the space/time coordinate  without error, where is the location in space and is the location in time. Note that this deterministic transformation has the useful property that at any s/t location of interest where is known, the corresponding air pollution concentration is given by the uniquely defined inverse transformation . With the de-trended log-transformed data defined by Eq. S2, we can then define as a zero-mean homogeneous/stationary space/time random field (S/TRF) having as one of its realizations. Conceptually a S/TRF can be defined as a random variable indexed by the s/t location , or alternatively as a collection of s/t realizations such that is one of these realizations. Hence the S/TRF captures the variability of the offset-removed log-concentration data . This variability depends on the air pollution data and the offset function that is chosen. In this work, we try two alternatives for . One is the separable s/t global offset (SSTO); the other is the composite s/t global offset (CSTO). We find that the latter works better for the California fire we are modeling.

The BME estimation method consists in processing general and site-specific knowledge available about in order to estimate its value across space and time. The general knowledge base (G-KB) that characterizes consists of its mean and the covariance functions. The mean function is , where  is the stochastic expectation. The covariance function describes ’s space/time dependencies between s/t locations ***p*** and ***p***’ and is expressed as

|  |  |  |
| --- | --- | --- |
|  |  | (S3) |

The site-specific knowledge base (S-KB) consists of hard data, *,* the offset-removed log-transformed observations located at s/t points *,* and soft data, *,* the offset-removed log-transformed model or satellite predictions located at s/t points . The uncertainty associated with the soft data is characterized in terms of a site-specific probability density function (PDF)  and corresponds to the centroids of the model or satellite grid cells.

In the BME framework, the G-KB is denoted as and the S-KB is denoted as *.* Using these definitions, we can summarize BME into three steps:

1. Examine the G-KB, using the *Maximum Entropy* principle of information theory to create the prior PDF , where is the value of at points  and is any estimation point of interest. Since the G-KB only includes statistical moments up to order two, then is the multivariable Gaussian PDF with means and covariance specified in .
2. Integrate the S-KB with the prior PDF, using an epistemic *Bayesian* conditionalization rule (Eq. S4 below), to create the BME posterior PDF, , providing a full stochastic representation of , the transformed air pollutant at the estimation point .
3. Compute the BME mean and corresponding posterior BME variance of the BME posterior PDF at estimation points on an estimation grid to obtain maps of the BME estimate and corresponding uncertainty of the offset-removed log transformed air pollutant. Then add back the offset and back log-transform in order to obtain maps of estimated air pollution concentrations.

The BME posterior PDF is given by the BME equation

|  |  |  |
| --- | --- | --- |
|  |  | (S4) |

where  is a normalization constant, is the offset-removed log-transformed observational data defined in Eq. S2, and is the PDF representing the uncertainty associated with the model or satellite soft data.

For this analysis, the BME framework was implemented in MATLAB R2017b using BMElib version 2.0c. The BME framework was implemented using an estimation neighborhood of 80 hard data points and 4 soft data points with a maximum temporal and spatial search radius of 7 days and 5 degrees. We used an estimation grid of 75 by 75 estimation points with hard data points and Voronoi points included. The maps produced are displayed at a 1-km resolution. Figure S5 provides a visual representation of how we implemented the BME estimation method using the hard data (observations), soft data (CC-CMAQ and Sat-PM2.5), and G-KB. As shown, at monitoring stations, the BME framework will estimate the concentration observed. The influence of each of these observations decreases with distance given the s/t covariance. At locations further from monitoring stations, the BME framework will trend towards estimating the CC-CMAQ/Sat-PM2.5 concentration if it has low associated uncertainty or will trend towards estimating the value of the global offset if the CC-CMAQ/Sat-PM2.5 concentration at that location has high associated uncertainty.

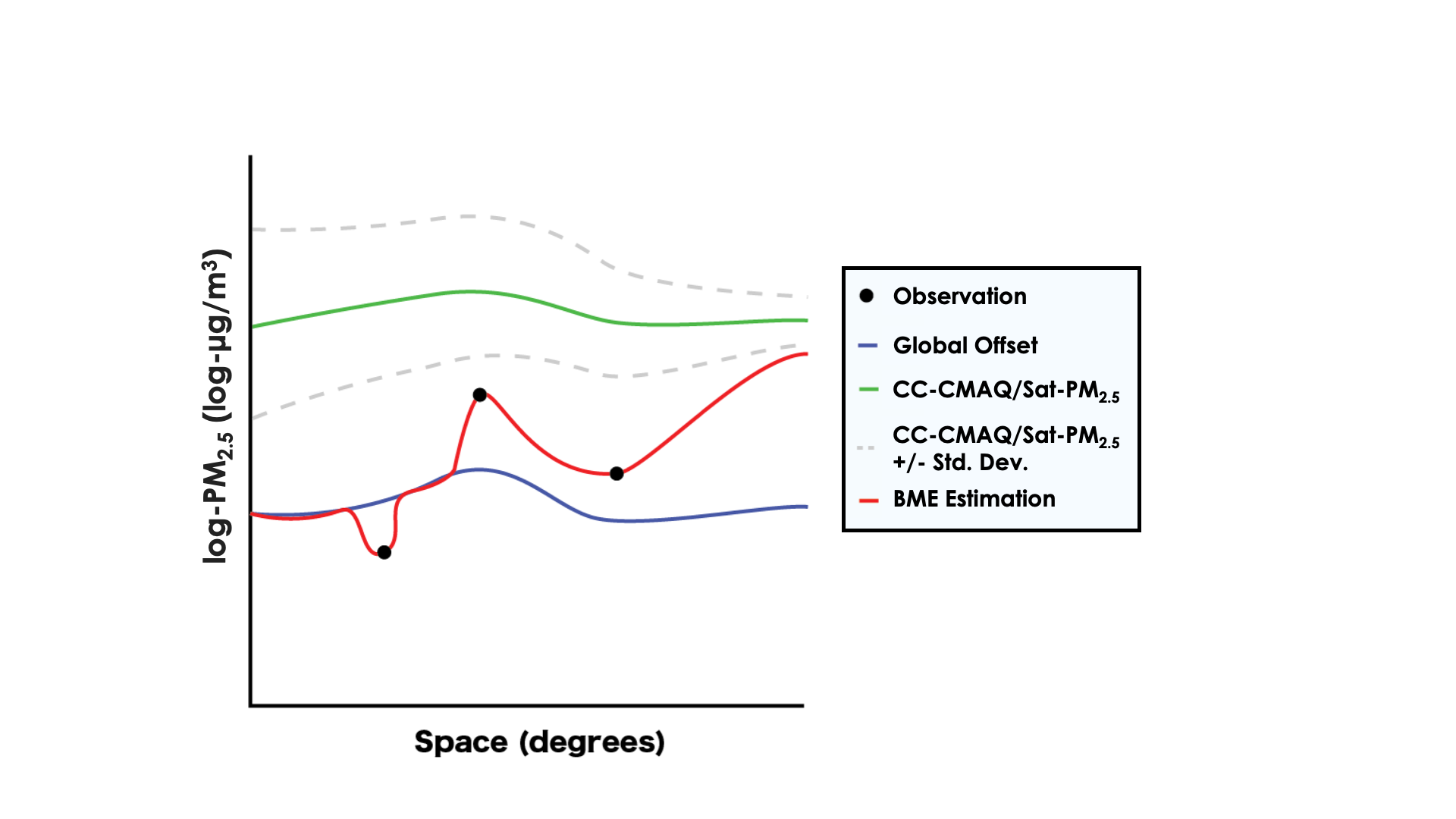


Figure S5. Visual representation of our implementation of the BME estimation method.

## S.5 Covariance of PM2.5 Data

The covariance model used in the G-KB to describe the s/t correlation of composite offset-removed natural log-PM2.5 values across California during October 2017 is defined

|  |  |  |
| --- | --- | --- |
|  |  | (S5) |

where r is the spatial lag (degrees) and is the temporal lag (days). The above equation is the standard format for covariance functions in the BME framework. The covariance parameters, derived from the offset-removed log-PM2.5 observations, are: ar1 = 0.15 degrees, at1 = 16,425 days, c01 = 0.0636 (log-µg/m3)2, ar2 = 4 degrees, at2 =365 days, c02 = 0.0142 (log-µg/m3)2, and ar3 = 2.5 degrees, at3 = 5 days, c03 = 0.441 (log-µg/m3)2*.* Figure S6 shows the fit of the covariance model to the data.

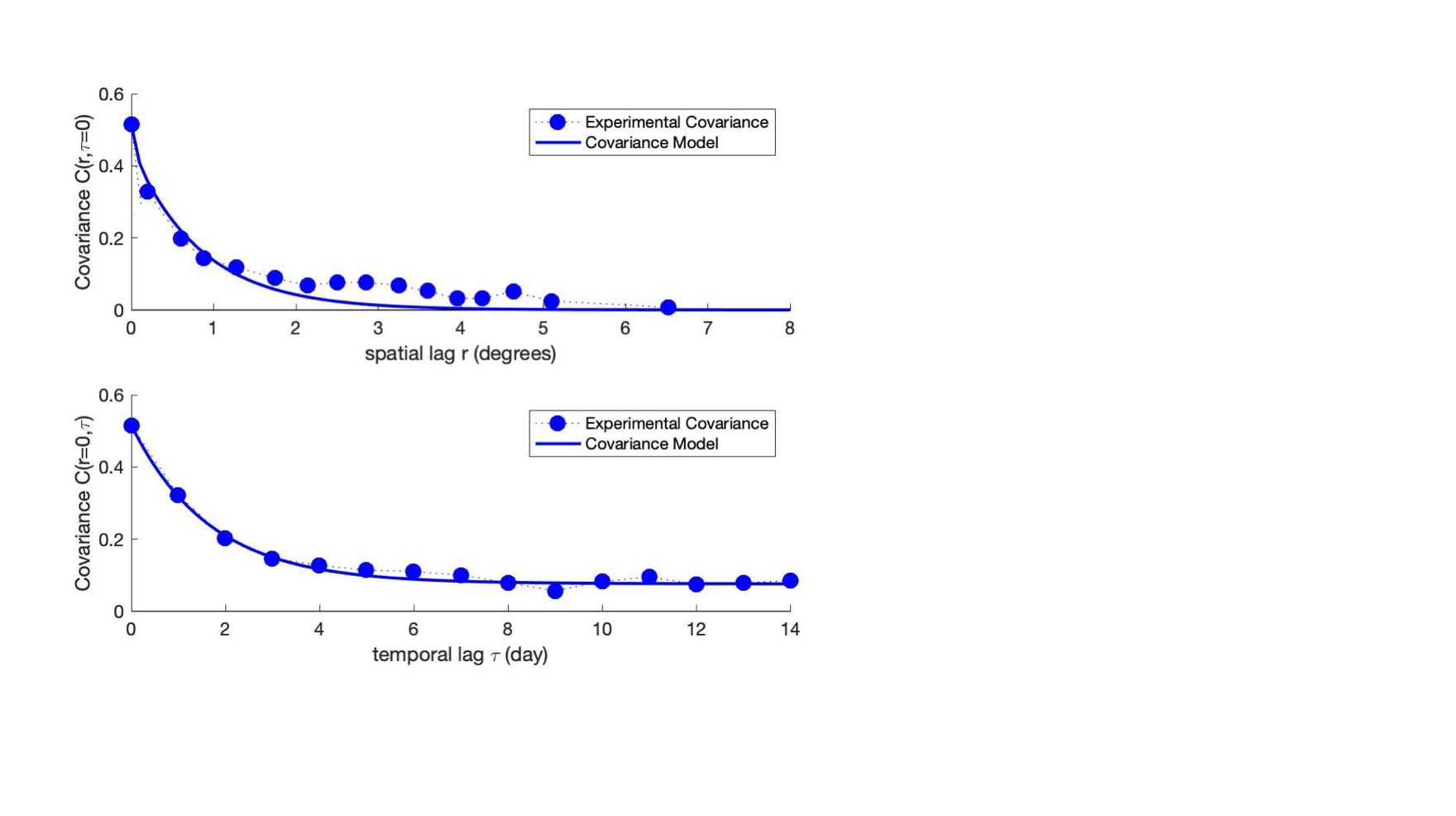


Figure S6. The fit of the covariance model to the experimental covariance values for both the spatial and temporal components for offset-removed log-PM2.5 levels across California during October 2017 wildfires.

This covariance model represents how the covariance for PM2.5 values in California during the October 2017 wildfires is the nesting of three covariance structures, all exponential in space and time. Each structure corresponds to factors that impact PM2.5 levels: human activity, weather events, and the October 2017 wildfires. The first two covariance structures are based on published covariance models of PM. Research shows that the covariance for annual mean PM is made up of two structures, both exponential in space and time, one structure of short temporal range and long spatial range, 1 year and 4 degrees (~444 km), that accounts for 15% of the variability and corresponds to fluctuations that are weather related, and one structure of long temporal range and short spatial range, 45 years and 0.15 degrees (~17 km), that accounts for 85% of the covariance and corresponds to fluctuations that are caused by long-term human activities in urban centers with high car traffic2. These exact parameters were used for the first two structures of the above covariance model, describing 15% of the total covariance, 12.75% for human activity and 2.25% for weather-related events. The other 85% of the covariance is described by the third, new structure, which corresponds to the October 2017 wildfires. With an intermediate spatial range of 2.5 degrees (~278 km) and a short temporal range of 5 days, the new structure can be explained by the size and duration of the October 2017 fires. This covariance model confirms that while the underlying causes of PM pollution were still present, the wildfires dominated how the PM2.5 levels covaried in California during October 2017.

## S.6 Additional Details on Global Offsets

The SSTO that we removed from the log-PM2.5 data is shown in Figure S7. The CSTO we removed from the log-PM2.5 data is shown in Figure S8 and S9. For the exponential smoothing function, both global offsets used the same parameters: a spatial neighborhood with a 2.5 degree radius, a temporal neighborhood with a 5 day radius, a 2 degree spatial smoothing range, and 5 day temporal smoothing range.

To construct the SSTO, we first obtain the raw temporal global offset (circles in Figure S7(1)) and the raw spatial global offset (colored circles in figure S7(2)), and we smooth them using the aforementioned exponential kernel smoothing parameters to obtain the smooth temporal global offset (line in Figure S7(1)) and the smooth spatial global offset (Figure S7(3)). In the SSTO, the equations for and are

|  |  |  |
| --- | --- | --- |
|  |  | (S6) |
|  |  | (S7) |

where is the number of PM2.5 observations within a set spatial radius of spatial location , is the number of PM2.5 observations within a set temporal radius of temporal location , and are the average of measurements at spatial location or temporal location , respectively, and and are the weights assigned to the average of measurements at that spatial or temporal location. and are determined by the spatial or temporal distance between and or and as well as the spatial or temporal ranges of the exponential smoothing function, described above.

Figure S9 emphasizes the differences between the SSTO and CSTO, highlighting the CSTO’s ability to capture different temporal trends at unique geographic locations.

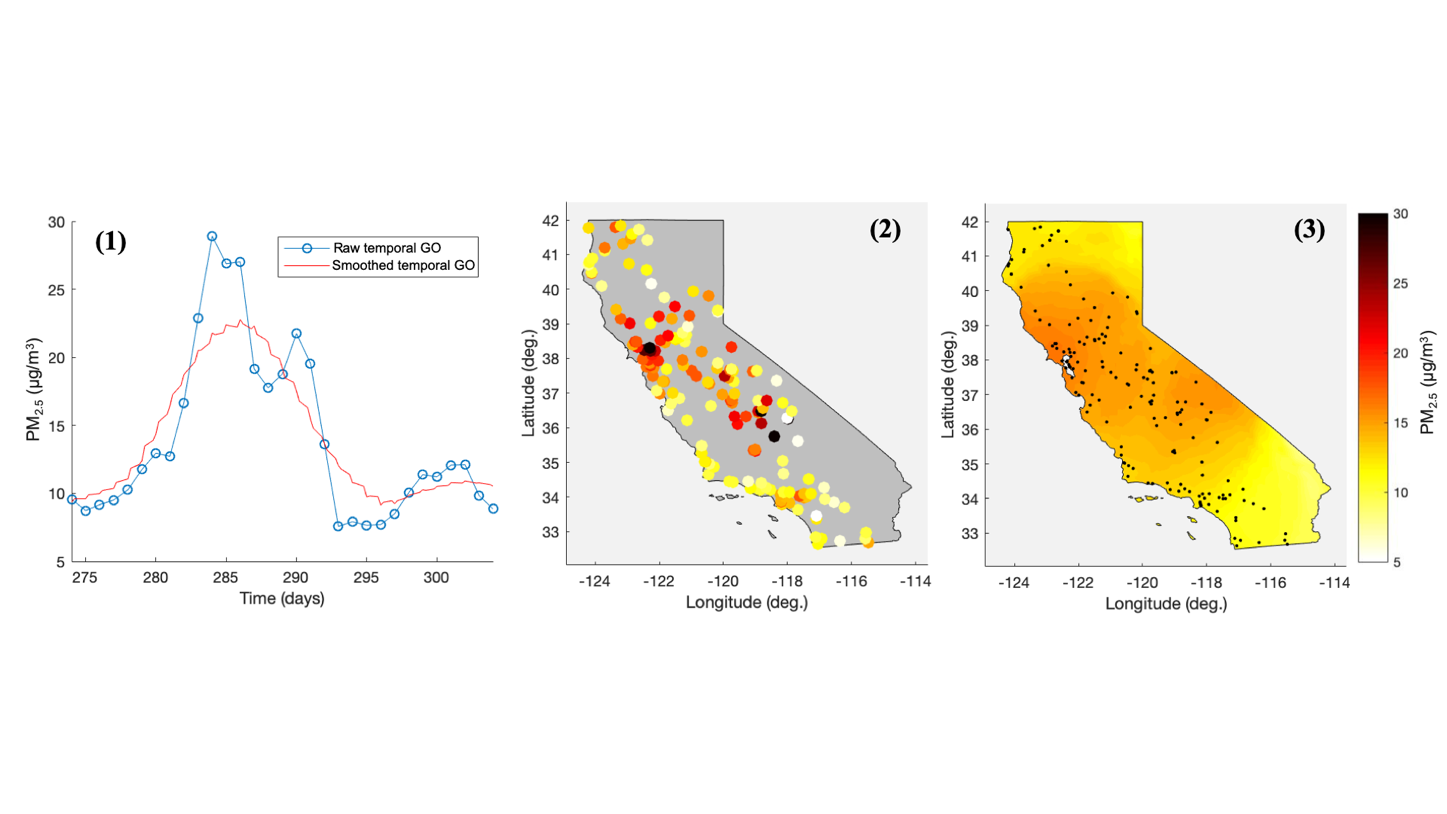


Figure S7. (1) The raw and smoothed temporal global offset, (2) the raw spatial global offset, and (3) smoothed spatial global offset for the SSTO that was removed from the log-PM2.5 data in California in October 2017.

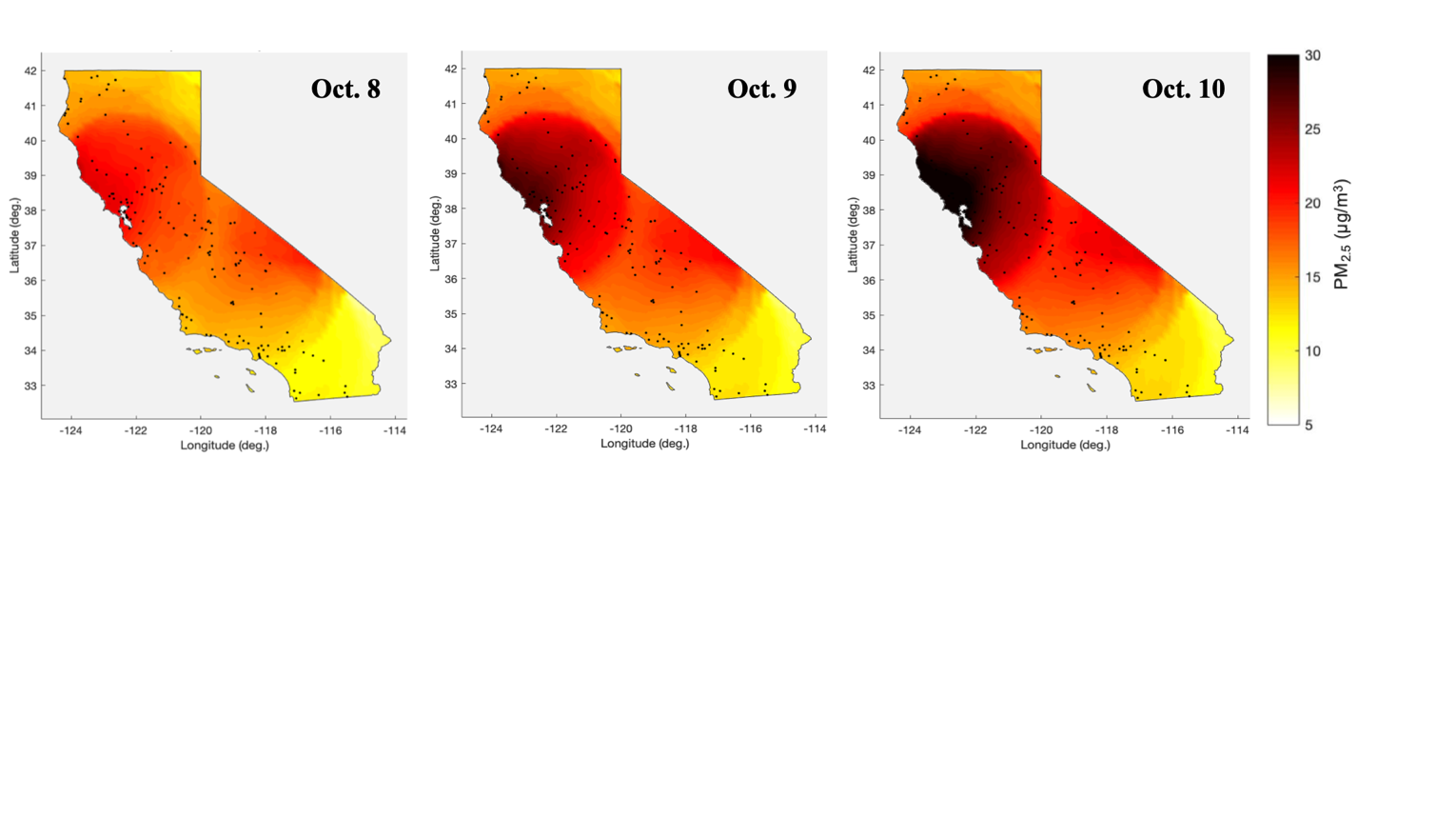


Figure S8. The spatial global offset, shown for Oct. 8-10, for the CSTO that was removed from the log-PM2.5 data in California in October 2017.



Figure S9. A comparison of the SSTO and CSTO smoothed temporal global offsets, shown at three different monitoring station locations across California in October 2017.

## S.7 Equations Used for Performance Evaluation

Table S1 provides equations for the performance statistics used in the LOOCV and RCV to assess the accuracy of the estimation of log-PM2.5 data by different BME methods. In Table S1, denotes a BME estimated value of log-PM2.5, or modeled or satellite-derived estimated value, at space/time point , is its paired observed value (i.e. observed at a monitoring station at the same space/time location) and is the corresponding error.

Table S1. Equations used to calculate cross-validation performance statistics.

|  |  |  |
| --- | --- | --- |
| **Metric Name** | **Definition** |  |
| # of data pairs |  | (S8) |
| Mean observed value |  | (S9) |
| Mean estimated value |  | (S10) |
| Variance observed value |  | (S11) |
| Variance estimated value (VZ) |  | (S12) |
| Mean error (ME) |  | (S13) |
| Variance of error (VE) |  | (S14) |
| Correlation () |  | (S15) |
| Correlation squared (R2) |  | (S16) |
| Mean squared error (MSE) |  | (S17) |

## S.8 State-Wide All-Month Performance Statistics

Table S2 shows the state-wide, all-month LOOCV performance statistics for the three BME s/t kriging approaches. Aligning with the LOOCV results in the fire-affected region and period, BME s/t kriging of both FRM/FEM and temporary station data with a CSTO provides the most accurate estimate across California between October 1-31. Adding temporary station data to the BME s/t kriging estimation increases accuracy, with a 44% reduction in MSE and a 36% increase in R2. Using a CSTO instead of a SSTO also improves estimation accuracy, with a 3% reduction in MSE and a 1% increase in R2.

**Table S2.** Leave-one-out cross-validation results for the estimation of the log of PM2.5 daily average concentrations across California between Oct. 1-31, using three BME s/t kriging approaches.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *BME s/t Kriging Method* | | **MSE \***  **(log-µg/m3)2** | **R2 \***  **(log-space)** | **ME \***  **(log-µg/m3)** | **VE \***  **(log-µg/m3)2** | **VZ \***  **(log-µg/m3)2** |
| **Observations** | **Global Offset** |
| FRM/FEM | Composite | 0.249 | 0.546 | 0.042 | 0.247 | 0.391 |
| FRM/FEM & Temporary | Separable | 0.144 | 0.735 | 0.001 | 0.144 | 0.472 |
| FRM/FEM & Temporary | Composite | 0.139 | 0.740 | 0.005 | 0.139 | 0.442 |

\* Performance metrics for the estimation of log-PM2.5 include mean square error (MSE), R-squared (R2), mean error (ME), variance of error (VE), and variance of estimation (VZ). The mean and variance of the observed log-PM2.5 data are 2.36 log-µg/m3 and 0.532 (log-µg/m3)2, respectively.

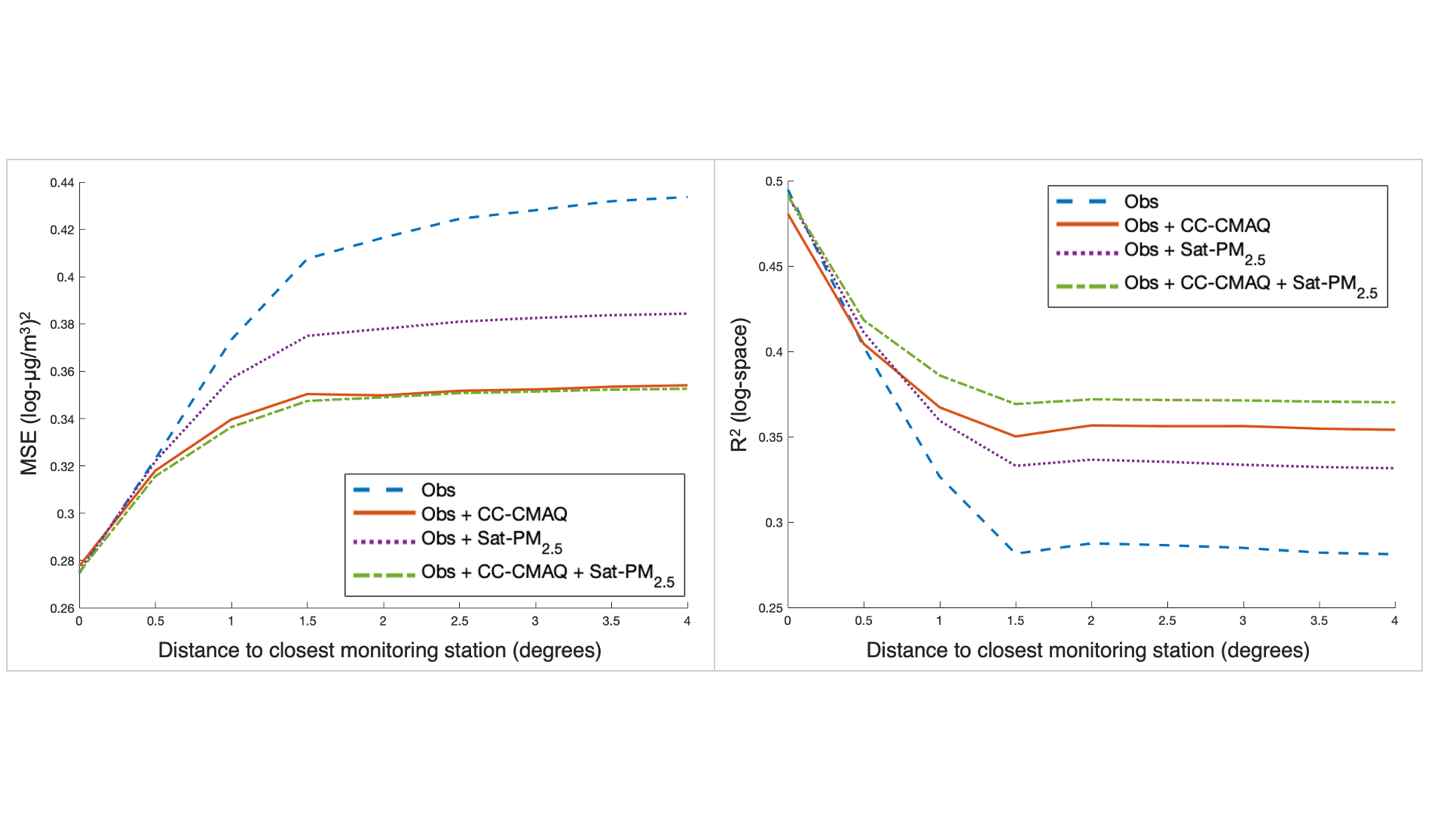
Table S3 shows the state-wide, all-month LOOCV performance statistics for the CAMP correction and the four BME methods. Aligning with the LOOCV results in the fire-affected region and period, bias-correcting the CMAQ output via CAMP notably improves performance across California between October 1-31, with a 57% decrease in MSE and a 21% increase in R2. Comparing the four BME approaches, BME s/t kriging of observations performs best across California during the month of October, with the lowest MSE and highest R2. All three BME data fusion approaches perform similarly but slightly worse in comparison. Overall, the state-wide, all-month LOOCV performance statistics are slightly better than those in the fire-affected region and period, with lower MSEs and higher R2s across all methods. The difference between the state-wide all-month performance and the fire-affected region and period performance is likely influenced by estimation accuracy in areas and on days outside the fire-affected region and period.

**Table S3.** Leave-one-out cross-validation results for the estimation of the log of PM2.5 daily average concentrations across California between Oct. 1-31, using CAMP correction, BME s/t kriging, and BME data fusion approaches.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **MSE \***  **(log-µg/m3)2** | **R2 \***  **(log-space)** | **ME \***  **(log-µg/m3)** | **VE \***  **(log-µg/m3)2** | **VZ \***  **(log-µg/m3)2** |
| Satellite-derived log-PM2.5 (Sat-PM2.5) | 0.339 | 0.367 | 0.147 | 0.317 | 0.180 |
| CMAQ Model | 0.703 | 0.410 | 0.178 | 0.672 | 1.235 |
| CAMP-Corrected (CC)-CMAQ Model | 0.304 | 0.496 | -0.003 | 0.304 | 0.315 |
| BME S/T Kriging | 0.139 | 0.740 | 0.005 | 0.139 | 0.442 |
| BME Data Fusion, Observations & CC-CMAQ | 0.145 | 0.727 | -0.007 | 0.145 | 0.380 |
| BME Data Fusion, Observations & Sat-PM2.5 | 0.144 | 0.730 | 0.016 | 0.144 | 0.375 |
| BME Data Fusion, Observations, CC-CMAQ, & Sat-PM2.5 | 0.147 | 0.726 | 0.010 | 0.147 | 0.338 |

\* Performance metrics for the estimation of log-PM2.5 include mean square error (MSE), R-squared (R2), mean error (ME), variance of error (VE), and variance of estimation (VZ). The mean and variance of the observed log-PM2.5 data are 2.36 log-µg/m3 and 0.532 (log-µg/m3)2, respectively.

Figure S10 shows the state-wide, all-month RCV results for the four BME approaches. Within 0.5 degrees of the closest station, all four BME approaches perform very similarly across California between October 1-31. Once an estimation location is more than 0.5 degrees from the closest station, the BME data fusion of observations with CC-CMAQ and Sat-PM2.5 performs best, with the lowest MSE and highest R2 at greater distances from the nearest station, aligning with the RCV results in the fire-affected region and period. The BME data fusion of observations with CC-CMAQ performs similarly, but slightly worse, in comparison and BME s/t kriging of observations performs worst. Overall, the state-wide, all-month RCV performance statistics are slightly worse than those in the fire-impacted region and period, with lower R2s across all three BME data fusion methods.



**Figure S10.** Results of radius cross validation for log-PM2.5 across California between Oct. 1-31; MSE (left) and R2 (right), based on distance from the closest monitoring station, for the 4 BME methods: BME s/t kriging, Observations (Obs); BME Data Fusion, Observations and CC-CMAQ (Obs + CC-CMAQ); BME Data Fusion, Observations and Sat-PM2.5 (Obs + Sat-PM2.5); BME Data Fusion, Observations, CC-CMAQ, and Sat-PM2.5 (Obs + CC-CMAQ + Sat-PM2.5).

## S.9 Refinement of Plume Shape with Addition of Temporary Data

Figure S11 provides a visual comparison of the BME s/t kriging PM2.5 estimations when only observations from FRM/FEM monitoring stations are used compared to when observations from both FRM/FEM and temporary monitoring stations are used. When the temporary station data is included in the BME estimation, there is an increase in observations in smoke-impacted regions, which results in a refinement of the smoke plume shape in northern California.

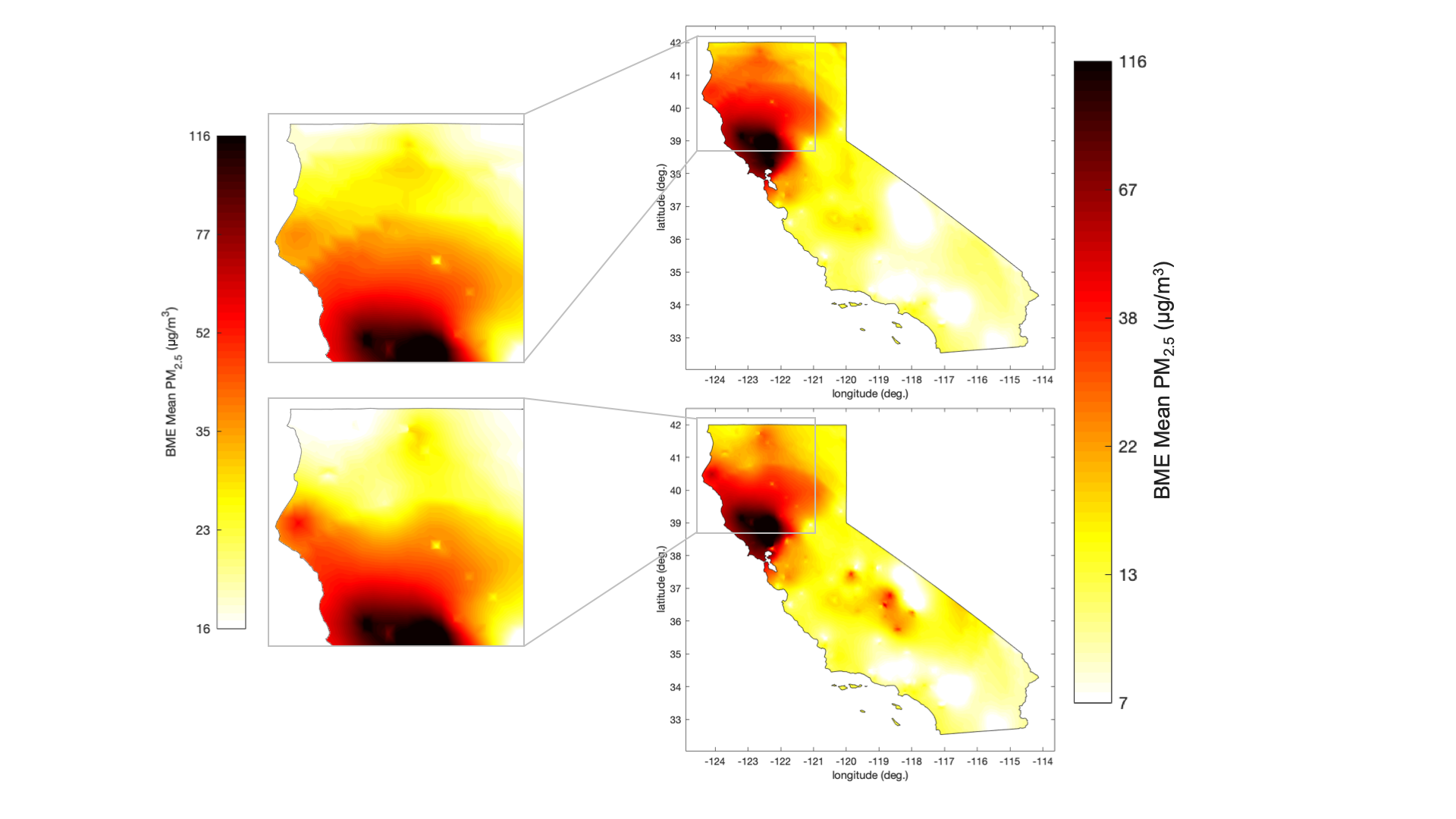


Figure S11. A comparison of two BME s/t kriging methods on Oct. 10, 2017, when (1) only observations from FRM/FEM monitoring stations are used and (2) observations from both FRM/FEM and temporary monitoring stations are used.

## S.10 Stratified Performance Statistics for Addition of Temporary Data

Table S4 shows the LOOCV performance statistics for the addition of temporary station data to the BME s/t kriging estimation in the fire-affected region and period, with the results stratified by station type. As shown in the Results section, when estimation performance is evaluated at both FRM/FEM and temporary monitoring station locations, the addition of temporary station observations results in a 36% increase in R2. When estimating at only the FRM/FEM station locations, the addition of temporary station observations to the BME s/t kriging estimation does not impact performance, with only a 0.1% reduction in R2. On the other hand, when estimating at only the temporary station locations, the addition of temporary station observations to the BME s/t kriging estimation notably improves performance, with a 44% increase in R2. The addition of temporary station data does not impact performance at FRM/FEM stations because of the existing availability of reliable information at these locations. The addition of temporary station observations proves most valuable in locations without FRM/FEM monitoring stations and improves the overall performance by increasing the availability of accurate information in otherwise station-scarce locations.

**Table S4.** Leave-one-out cross-validation results for the BME s/t kriging estimation of the log of PM2.5 daily average concentrations in the fire-affected region and period, with and without observations from temporary stations. Results are stratified by station type.

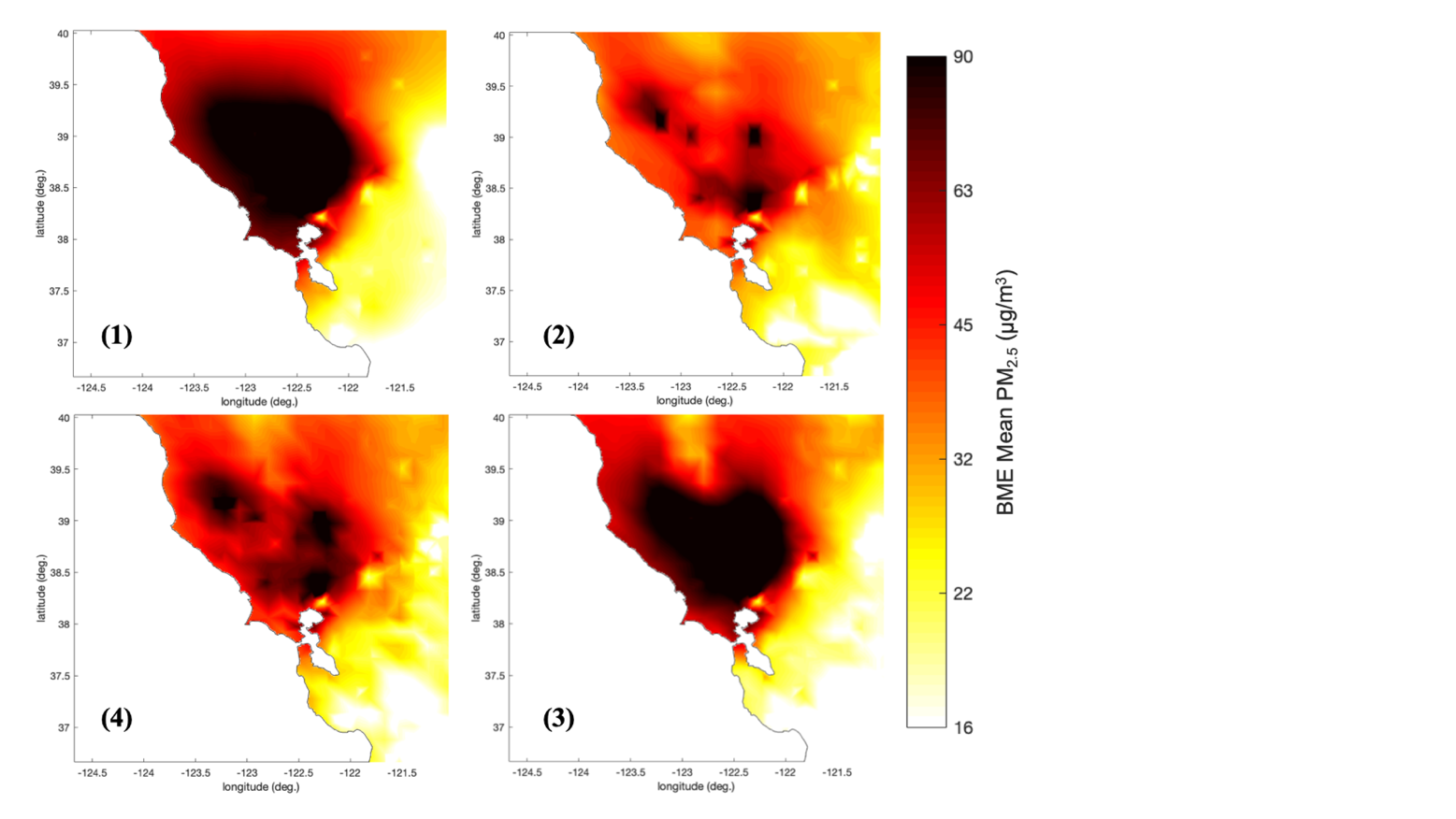
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | *FRM/FEM & Temporary* | | *FRM/FEM* | | *Temporary* | |
| *BME s/t Kriging Method* | | **MSE \***  **(log-**  **µg/m3)2** | **R2 \***  **(log-space)** | **MSE \***  **(log-µg/m3)2** | **R2 \***  **(log-space)** | **MSE \***  **(log-µg/m3)2** | **R2 \***  **(log-space)** |
| **Observations** | **Global Offset** |
| FRM/FEM | Composite | 0.327 | 0.520 | 0.120 | 0.814 | 0.409 | 0.284 |
| FRM/FEM & Temporary | Composite | 0.196 | 0.708 | 0.132 | 0.815 | 0.302 | 0.507 |

\* Performance metrics for the estimation of log-PM2.5 include mean square error (MSE) and R-squared (R2). The mean and variance of the observed log-PM2.5 data are 2.36 log-µg/m3 and 0.532 (log-µg/m3)2, respectively.

## S.11 Further Comparison of BME Methods

Figure S12 shows a comparison of the 4 BME methods in the Bay Area, the most impacted region during the October 2017 fires. As shown, the addition of either CC-CMAQ or Sat-PM2.5 results in a noticeable refinement of the smoke plume shape. By combining both, the BME data fusion of observations with CC-CMAQ and Sat-PM2.5 incorporates valuable information from each output and as a result provides the most informed, physically meaningful estimate of the smoke plume shape and location.

Figure S13 shows the difference in PM2.5 estimation values between BME s/t kriging of observations and the BME data fusion of observations with CC-CMAQ and Sat-PM2.5. These difference maps, shown for the first six days of the October 2017 wildfires, emphasize that including CC-CMAQ and Sat-PM2.5 in the BME estimation impacts the estimated values of PM2.5 concentrations, especially in fire-impacted regions. Areas in blue show where incorporating CC-CMAQ and Sat-PM2.5 into the BME framework decreased the estimated PM2.5 concentration while areas in red show where incorporating CC-CMAQ and Sat-PM2.5 increased the estimated PM2.5 concentration. Areas in green show where including CC-CMAQ and Sat-PM2.5 had no impact on the BME estimation.



**Figure S12.** Comparison of 4 BME methods to calculate the median value of daily average PM2.5 concentrations in the Bay Area on Oct. 10, 2017. (1) BME s/t kriging, Observations; (2) BME Data Fusion, Observations and CC-CMAQ; (3) BME Data Fusion, Observations and Sat-PM2.5; (4) BME Data Fusion, Observations, CC-CMAQ, and Sat-PM2.5.

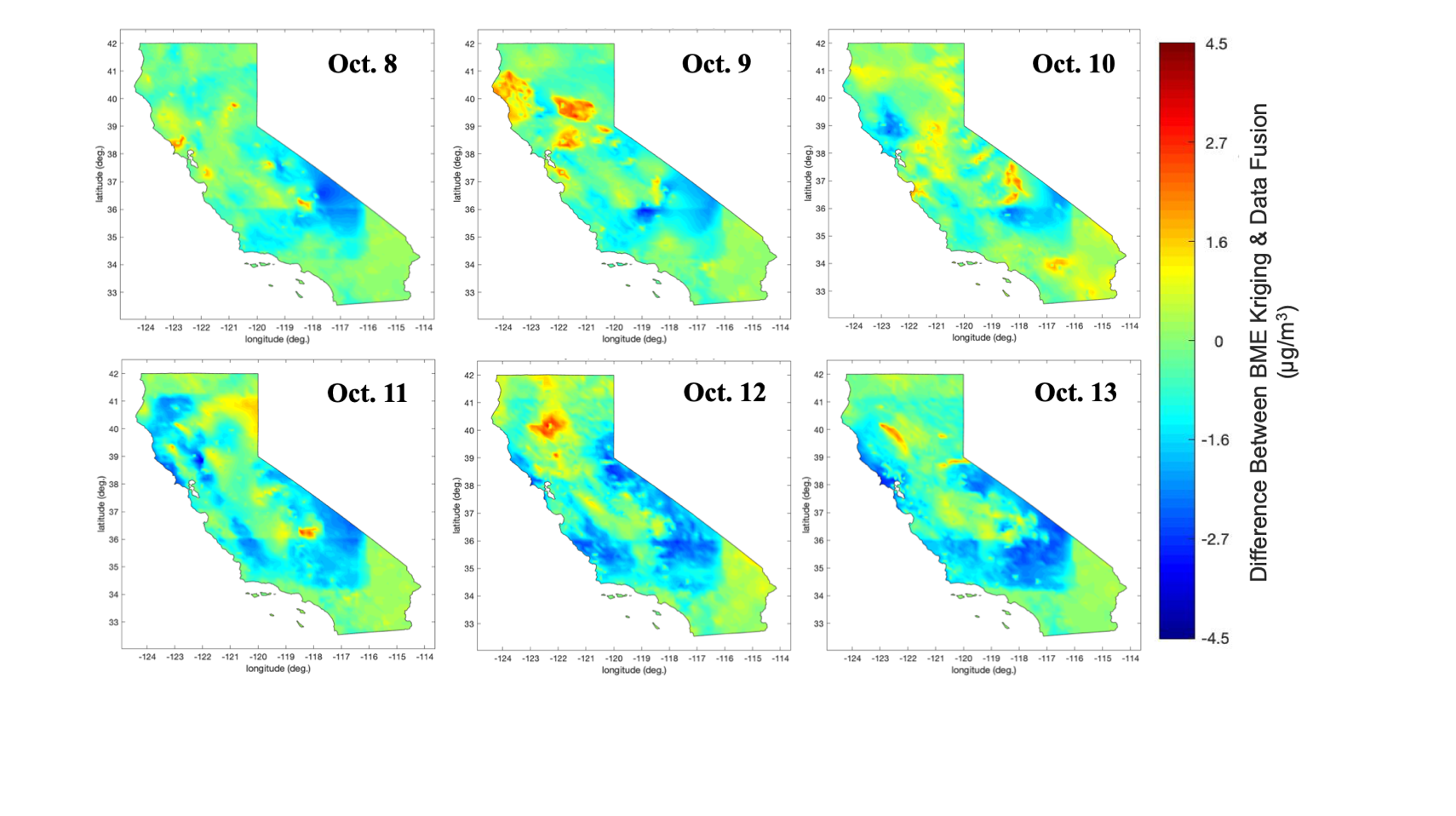


Figure S13. The difference in PM2.5 estimation values between BME s/t kriging of observations and BME data fusion of observations, CC-CMAQ, and Sat-PM2.5 for Oct. 8-13, 2017.

## S.12 Animation of Estimated PM2.5 Concentrations

We compiled a time-lapsed animation of PM2.5 estimations across California between October 8-20, 2017. These estimation maps were created through the BME data fusion of observations with CC-CMAQ and Sat-PM2.5. The animation may be viewed free of charge online at the following website: <https://mserre.sph.unc.edu/BMElab_web/mappingStudies/SC_PM25_CA_2017_Oct_08-20/>

A 1-km resolution CSV of PM2.5 estimations across California for October 8-20, 2017 is available upon request.

## S.13 Variance Maps of Estimated PM2.5 Concentrations

Figure S14 shows the BME variance for the log-PM2.5 concentration estimates produced through the BME data fusion of observations, CC-CMAQ, and Sat-PM2.5 for the first six days of the fires. The areas with lowest estimation variance occur at the monitoring station locations. At these points, the BME framework estimates the observed value, resulting in no associated uncertainty. The estimation variance increases with distance from these stations, given the s/t covariance of the observations. Within the CMAQ domain and further from monitoring stations, the variance is informed by both the uncertainty of CC-CMAQ, the value generated during CAMP correction process, and the uncertainty of Sat-PM2.5, the variance generated during the AOD to PM2.5 conversion. Outside the bounds of the CMAQ domain and further from monitoring stations, the estimation variance is informed solely by the uncertainty of Sat-PM2.5. In regions where there is no Sat-PM2.5 or monitoring station data, the estimation variance increases sharply with increased distance from monitoring stations. These estimation variance maps give confidence in the accuracy of the daily average estimation maps produced given the overall low variance, especially in northern California and the Bay Area, the regions most impacted by smoke during the October 2017 wildfires.

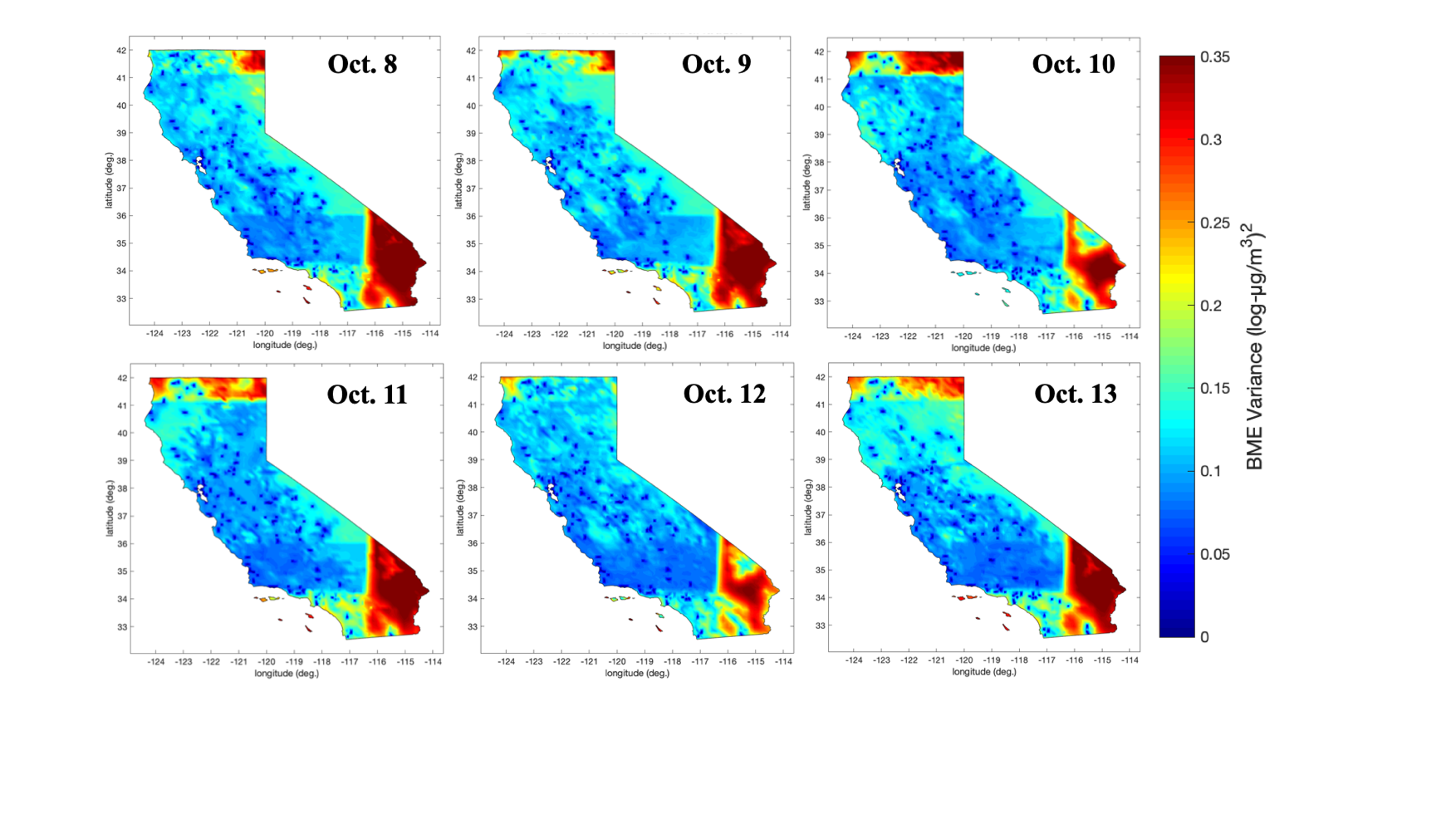


Figure S14. Variance of the estimated log of daily average ground-level PM2.5 concentrations, produced through the BME data fusion of observations, CC-CMAQ, and Sat-PM2.5, across California for Oct. 8-13, 2017.

## References

(1) Xu, Y.; Serre, M. L.; Reyes, J.; Vizuete, W. Bayesian Maximum Entropy Integration of Ozone Observations and Model Predictions: A National Application. *Environ. Sci. Technol.* **2016**, *50* (8), 4393–4400. https://doi.org/10.1021/acs.est.6b00096.

(2) Serre, M. L.; Christakos, G.; Lee, S. J. Soft Data Space/Time Mapping of Coarse Particulate Matter Annual Arithmetic Average Over the U.S. In *geoENV IV — Geostatistics for Environmental Applications*. **2006**, pp 115–126. https://doi.org/10.1007/1-4020-2115-1\_10.