Mapping the PM2.5 distribution and emerging hotspots in California for 2020

Ang Xiang Jing
Ho Shi Yun
Lim Zhu An
Tse Wing Yan

Introduction

- Exposure to fine particulate matter may result in the development of respiratory and cardiovascular diseases
- It is a threat to public health (Vodonos & Schwartz, 2021) and public and environmental welfare (U.S. EPA, 2012).

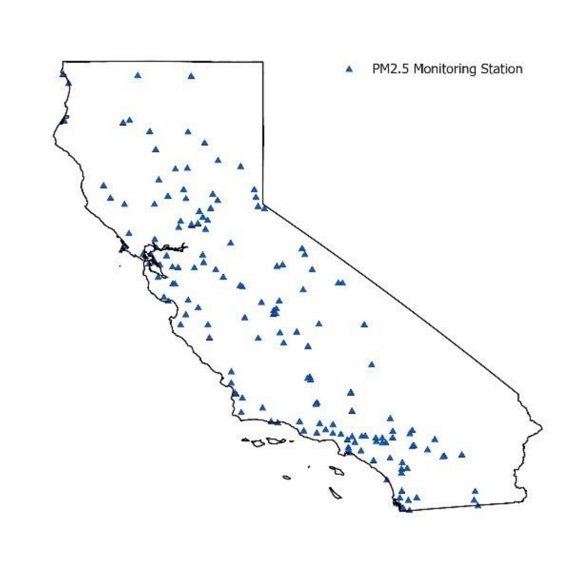
Source: LA Times - https://www.latimes.com/california/story/2020-09-10/air-quality-california-fires-bay-area-los-angeles

Objective & Goals

- Map the monthly distribution of PM2.5 in California;
- Perform map algebra operations to calculate the mean, minimum, maximum, and range of PM2.5 values for 2020;
- Identify space-time PM2.5 emerging hotspots and visualise them in 2D and 3D;
- Utilise census tract data to identify vulnerable populations who are at risk of being exposed to unhealthy PM2.5 levels.

Study Area

- We focused on California as it is the state with the worst air quality in the USA (Bashir et al., 2020)
- Significant correlation between environmental pollutants (e.g. PM2.5, PM 10, SO_2, CO and NO_2) for total cases and total mortality (Bashir et al., 2020)
- Exhibits high variability in PM2.5 sources, meteorology and topography (Li et al., 2020)



PM2.5 Air Quality Index

When the daily average PM2.5 value rises above 35.4 µg/m^3, it is considered **unhealthy** for sensitive groups.

AQI Category	Index Values	Previous Breakpoints (1999 AQI) (µg/m^3, 24-hour average)	Revised Breakpoints (µg/m^3, 24-hour average)
Good	0 - 50	0.0 - 15.0	0.0 – 12.0
Moderate	51 - 100	>15.0 - 40	12.1 – 35.4
Unhealthy for Sensitive Groups	101 – 150	>40 – 65	35.5 – 55.4
Unhealthy	151 – 200	> 65 – 150	55.5 – 150.4
Very Unhealthy	201 – 300	> 150 – 250	150.5 – 250.4
Hazardous	301 – 400	> 250 – 350	250.5 – 350.4
	401 – 500	> 350 – 500	350.5 – 500

Data Sources

S/N	Dataset Name	Dataset Description	Data Source	Data Format	Last Updated
1	California 2020 PM2.5 measurements	Daily mean PM2.5 measurements from fixed monitoring stations across California state in 2020.	US Environmental Protection Agency (EPA)	CSV	2021
2	OAFN California County Census	California county boundary with respective Census data requested by OAFN.	California Office of Access and Functional Needs (OAFN)	SHP	8/8/2017

Literature Review

PM 2.5 studies in California

Research direction

Mapping and visualisation

- Spatial-temporal dimension

(Our research focus), rarely examined in California

- Estimative and predictive modelling

Aguilera et al., 2020; Li et al., 2020; Stowell et al., 2020; Xiao et al., 2020

- Correlation between variables

Bashir et al., 2020; Rooney et al., 2020; Vodonos & Schwartz, 2021

Statistical methods chosen

- Spearman and Kendall correlation tests
- Inverse Distance Weighting (IDW)
- General Additive Model (GAM)

The effect of Wildfire to PM 2.5

Concur: (Li et al., 2020; Rooney et al., 2020; Stowell et al., 2020)

- Santa Ana seasonal winds exacerbates propensity for fire ignition and prolongs fire seasons (Stowell et al., 2020)

Reject: Cisneros et al. (2014)

Literature Review

Spatial-temporal mapping and spatial data handling

	Method	Scholar	Rationale / strength
Interpolation	**Inverse Distance Weighting (IDW)**	Aguilera et al., 2020; Keler & Krisp, 2015	1. Suitable when data points are sparse 2. Noted to be preferable and precise over the option of having estimated values to be within range of sample values
	Ordinary Kriging (OK)	Huang et al., 2015	• Shows a positive correlation between uncertainty and distance away from points (Brus et al., 2013)
	Empirical Bayesian Kriging (EBK)	Bhunia & Ding, 2020; Yang et al., 2017; Zhang et al., 2016; Brown et al., 1994	1. Its capability to interpolate spatially intensive data 2. Provide a dependable diagnosis of the uncertainty of model predictions Normal Quantile-Quantile plot • Study spatiotemporal linkages among air quality variables Performance evaluation • ArcGIS Pro allows the generation of Root Mean Squared Error (RMSE) and Standard Error Maps to check for interpolation methods' performance
	Geographically Weighted Regression (GWR) model	Luo et al., 2017; Yang et al., 2017	• Pays attention to spatial variabilities • Accounts for the dynamic nature of auxiliary variables in different areas and seasons
	Radial Basis Function (RBF)	Bhunia & Ding, 2020	
Visualisation	**Histogram**	Yang et al., 2017	• Plot against time to identify the highest concentration period
	Hotspot Polygon	Keler & Krisp, 2015	• Quick visualisation of outstanding values
	Generalized Additive Mixed model (GAMM)	Cisneros et al., 2014; Huang et al., 2015	• Specific for analyzing the linkages between PM 2.5 and other variables

Methodology and Workflow

Pre-processing of PM 2.5 data

- Removing of outliers (values <0 µg/m³)
- Filtered and export data for1st day of every month
- Reproject to EPSG 3311

Exploratory Data Analysis

- Using boxplot, histogram, QQ plot, summary statistics and attribute table functionality

Spatial Interpolation using Empirical Baysesian Kriging (EBK) + Cell Statistics

- Performed EBK with multiplicative skewing normal score transformation
- Mean, minimum, maximum and range of PM2.5 across 12 months were calculated

Create Space Time Cube

- Fishnet: 10 x 10km
- Extract interpolated PM2.5 to fishnet centroids

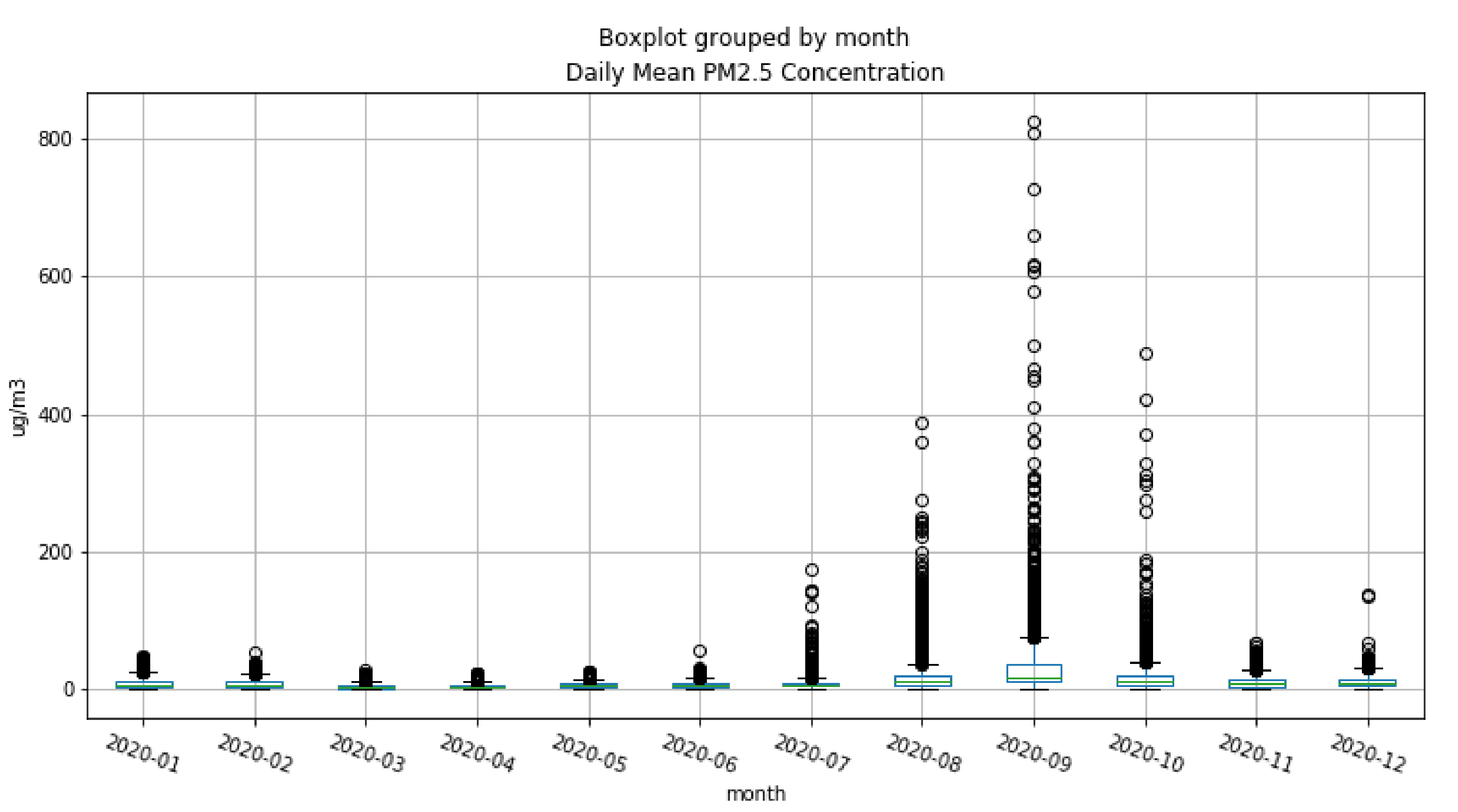
Perform Emerging Hotspot Analysis (EHSA)

2D & 3D visualisation of EHSA results

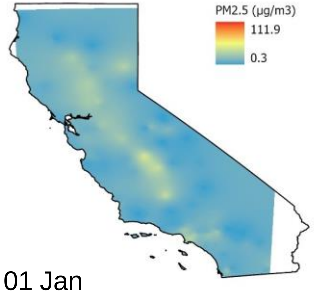
Indicator Kriging for Sep-Dec 2020

Calculate and rank vulnerability (for Oct 2020)

Results – Boxplot summary



Results – EBK Prediction Maps



01 Jan 2020

01 Feb 2020

01 Mar 2020

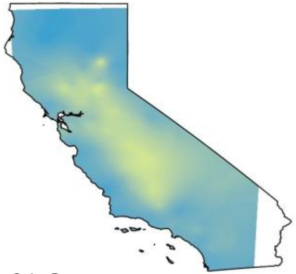
01 Apr 2020

01 May 2020

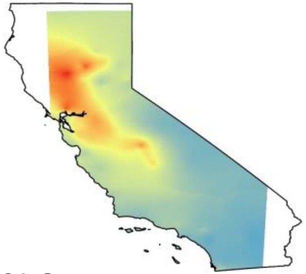
01 Jun 2020

01 Jul 2020

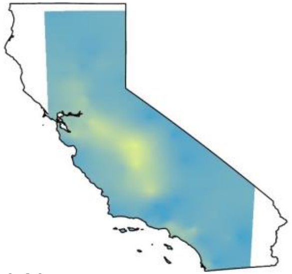
01 Aug 2020



01 Sep 2020



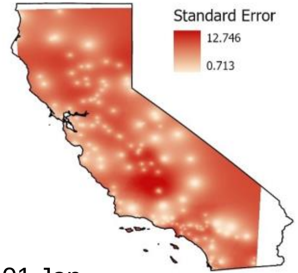
01 Oct 2020



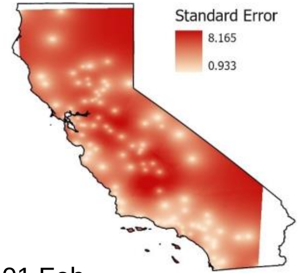
01 Nov 2020

01 Dec 2020

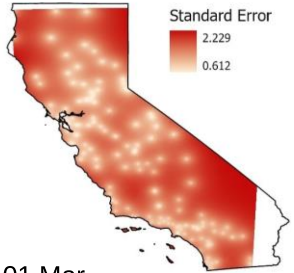
Results – EBK Standard Error Maps



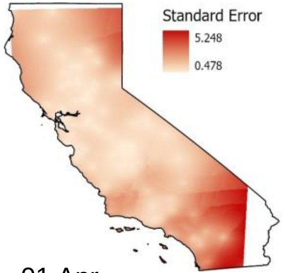
01 Jan 2020



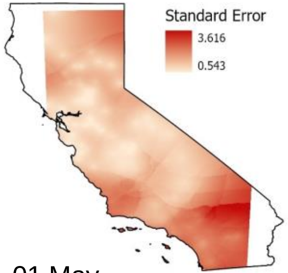
01 Feb 2020



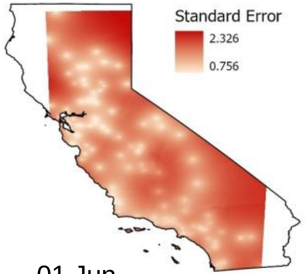
01 Mar 2020



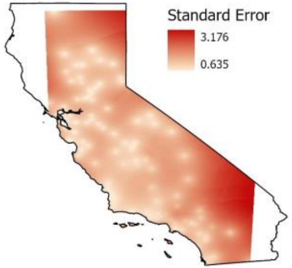
01 Apr 2020



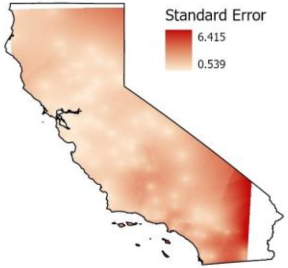
01 May 2020



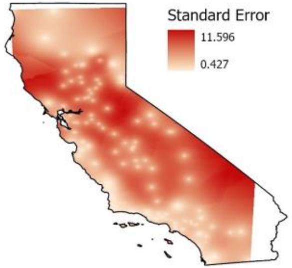
01 Jun 2020



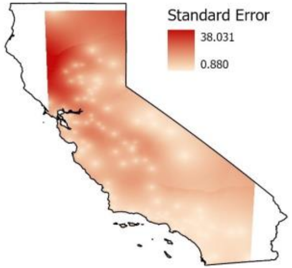
01 Jul 2020



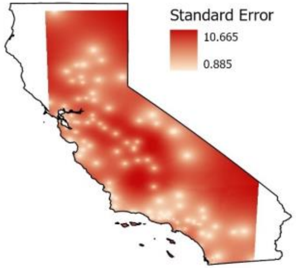
01 Aug 2020



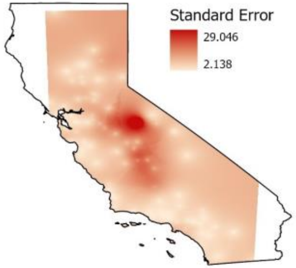
01 Sep 2020



01 Oct 2020



01 Nov 2020

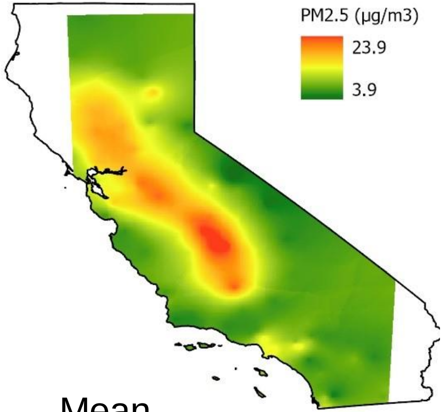


01 Dec 2020

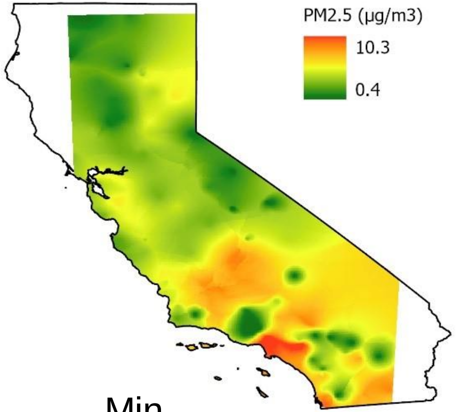
Results – Cross validation Summary Statistics

Month	RMSE	RMSE (Standardised)	Average SE
Jan-01	6.813	1.012	7.283
Feb-01	5.220	1.030	5.217
Mar-01	1.852	1.000	1.864
Apr-01	2.695	0.969	2.717
May-01	3.026	0.914	3.079
Jun-01	2.121	1.001	2.100
Jul-01	2.039	1.065	1.962
Aug-01	3.056	1.077	2.734
Sep-01	6.833	1.434	5.652
Oct-01	12.430	1.134	11.921
Nov-01	6.041	0.961	6.573
Dec-01	14.319	1.513	9.668

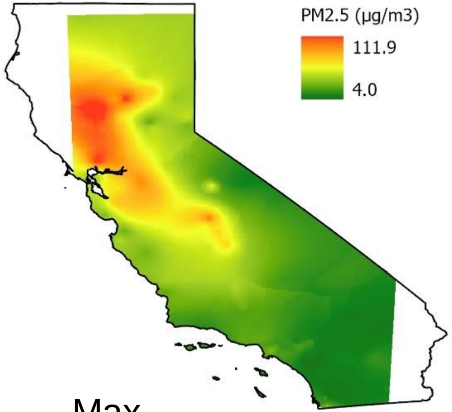
Results – Cell Statistics



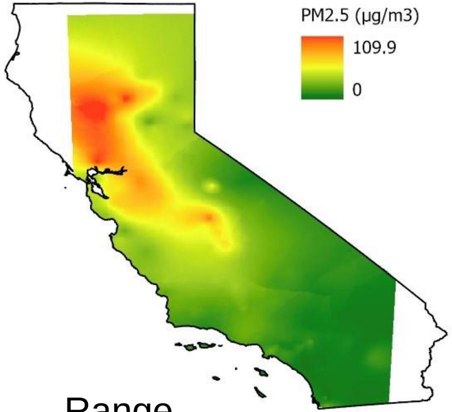
Mean



Min

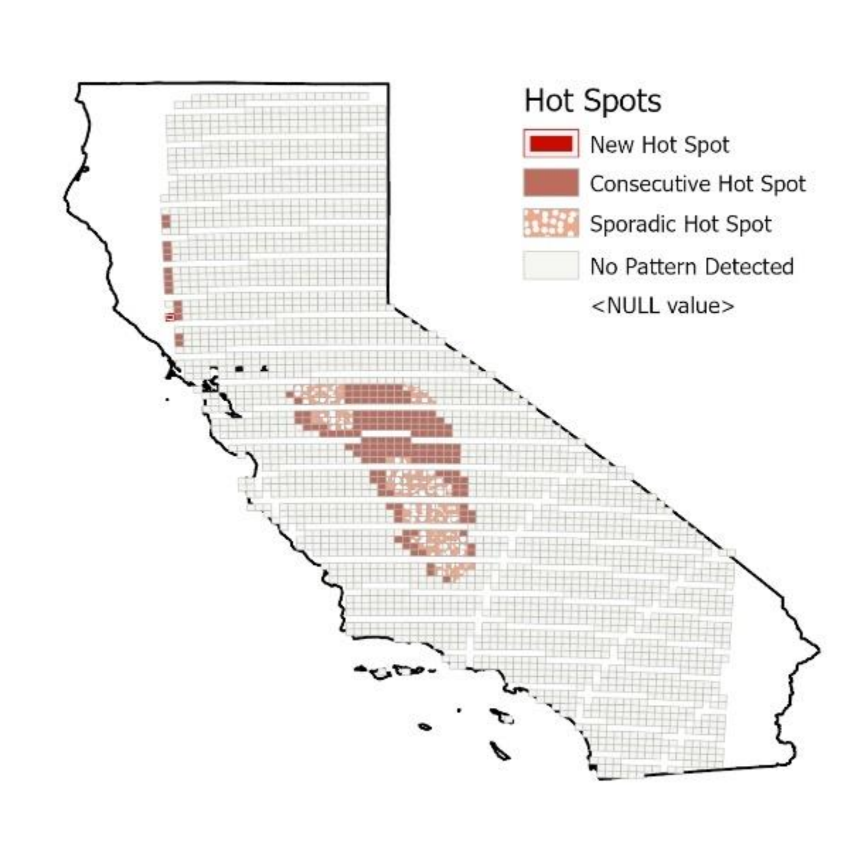


Max

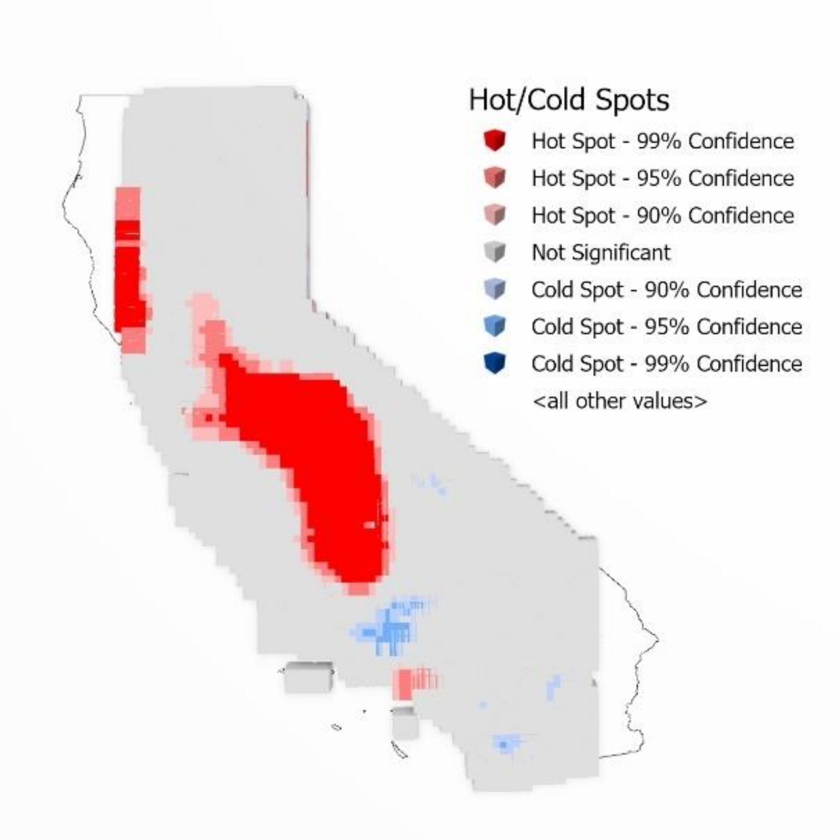


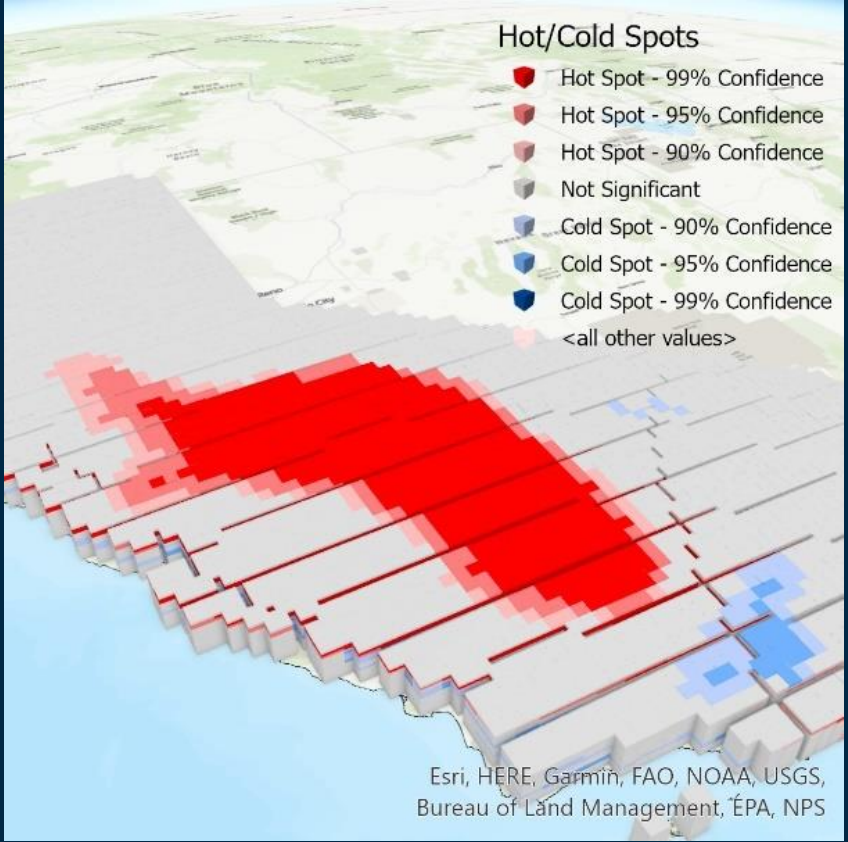
Range

Results – EHSA (2D)

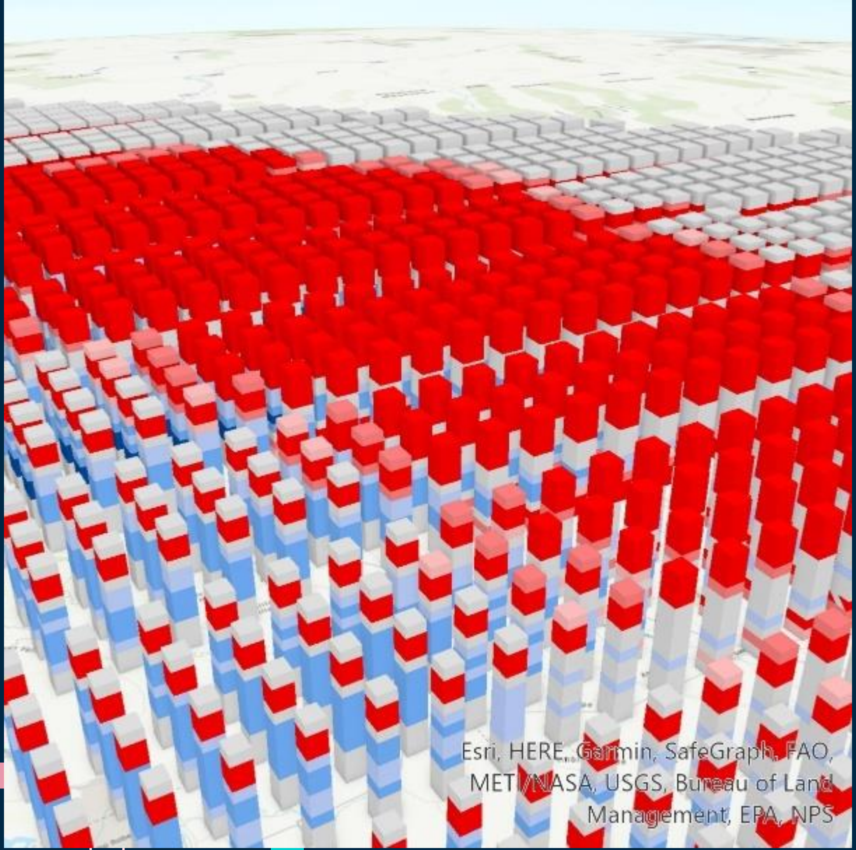


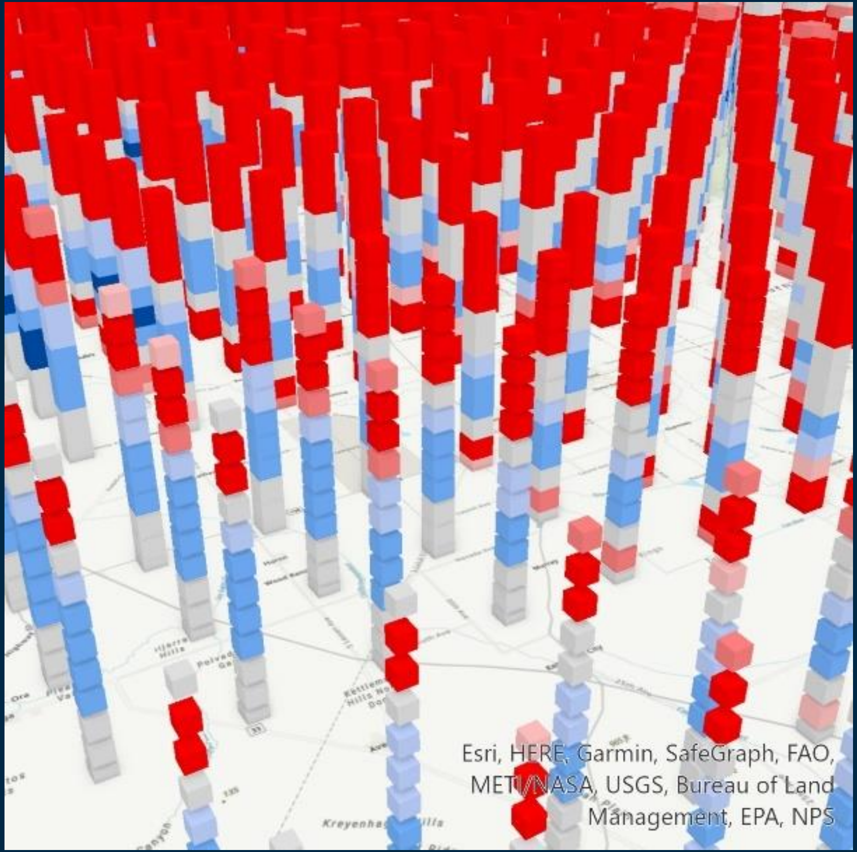
Results – EHSA (3D)



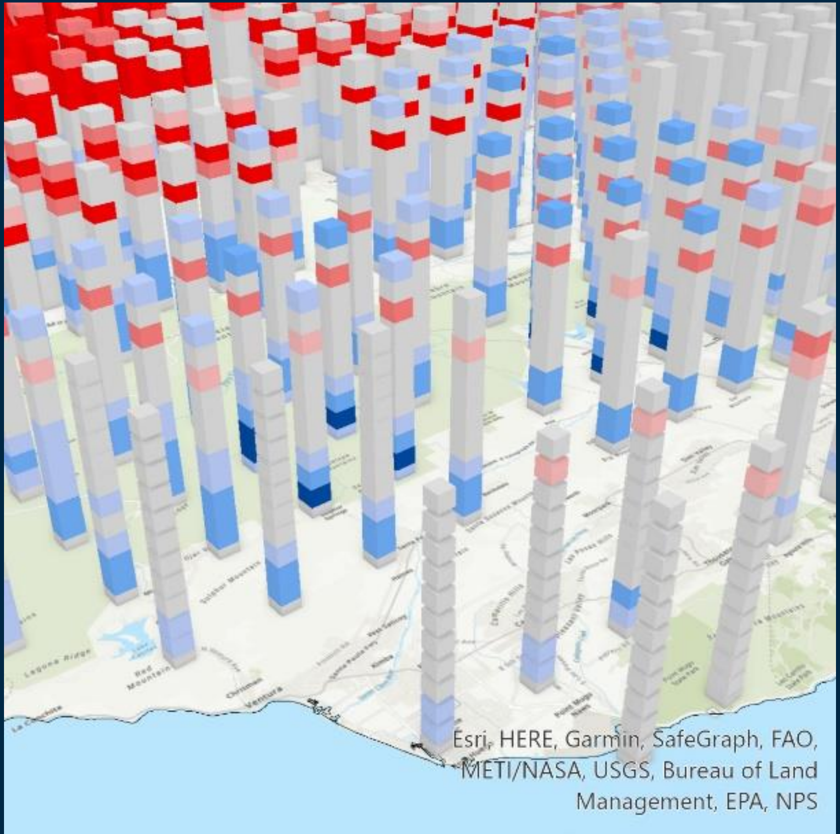


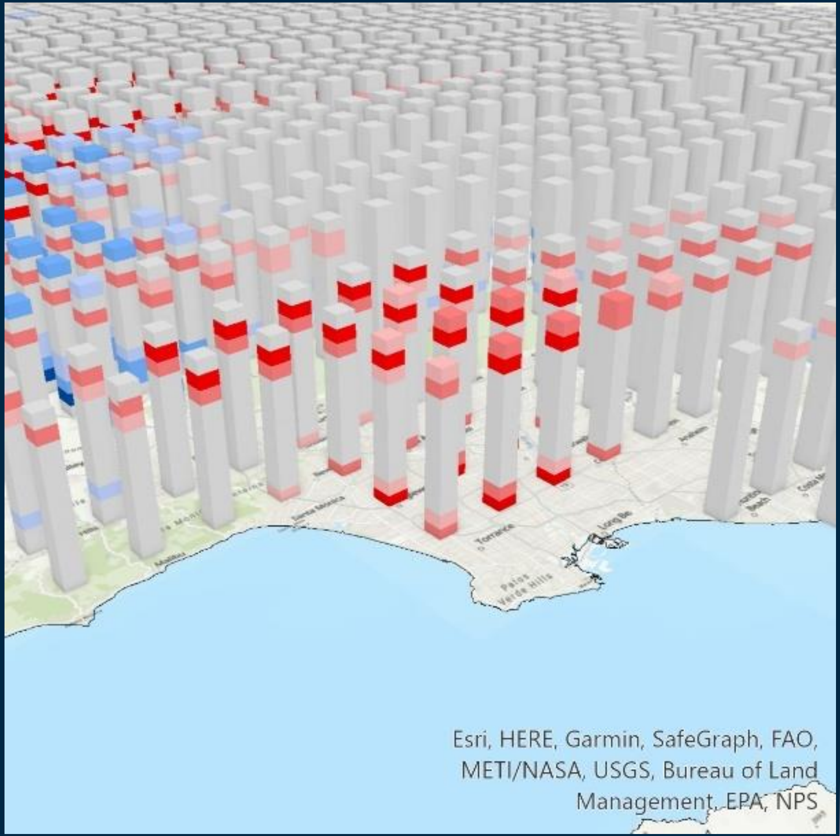
Results – EHSA (3D)



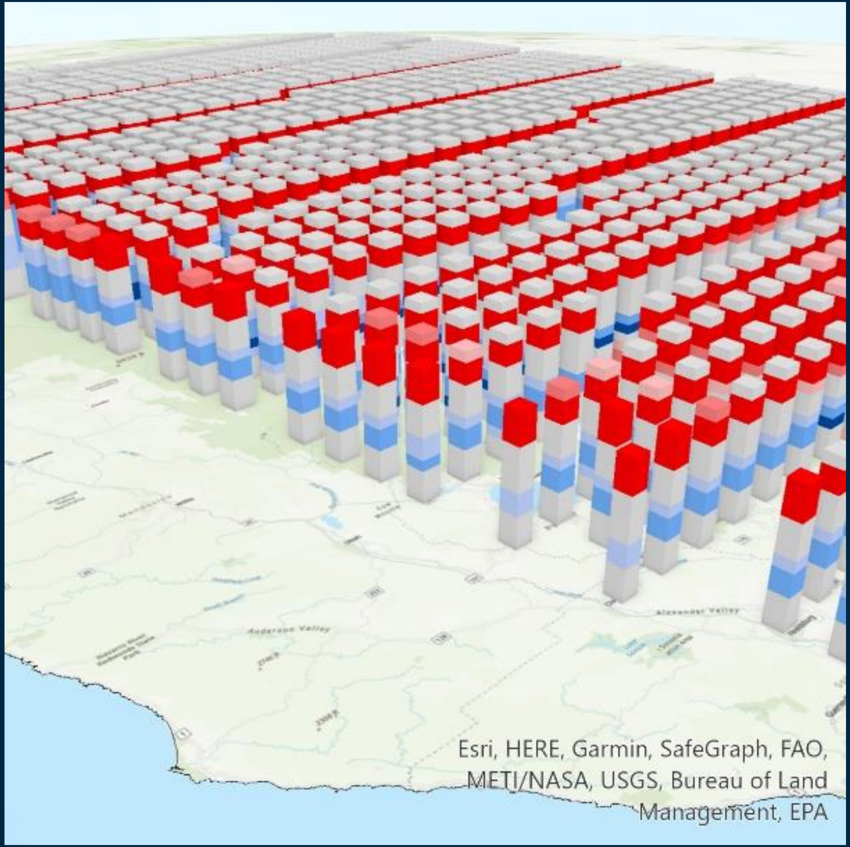


Results – EHSA (3D)





Results – EHSA (3D)



Results – Vulnerable Population & Calculation

- Vulnerable population are young children < 5 years old, and the higher age-group populations of ages 45-64 and >65 years old.

$$V_m = P * \frac{Pop_i - Pop_{min}}{Pop_{max} - Pop_{min}}$$

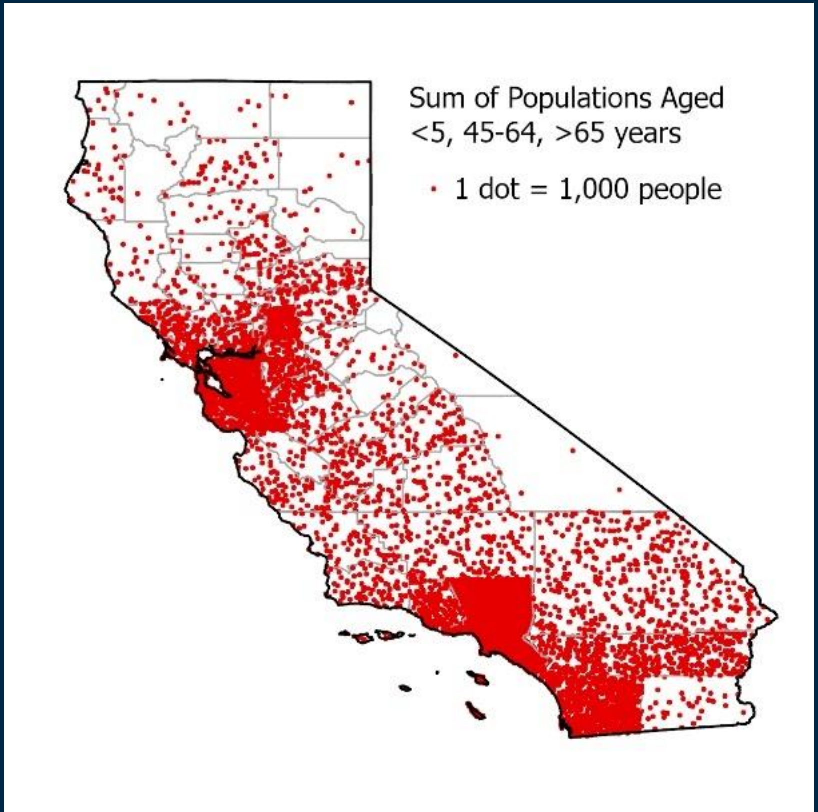
where V is the vulnerability for a given month m,

P is the probability that the unhealthy PM2.5 threshold for vulnerable groups will be exceeded (derived from Indicator Kriging),

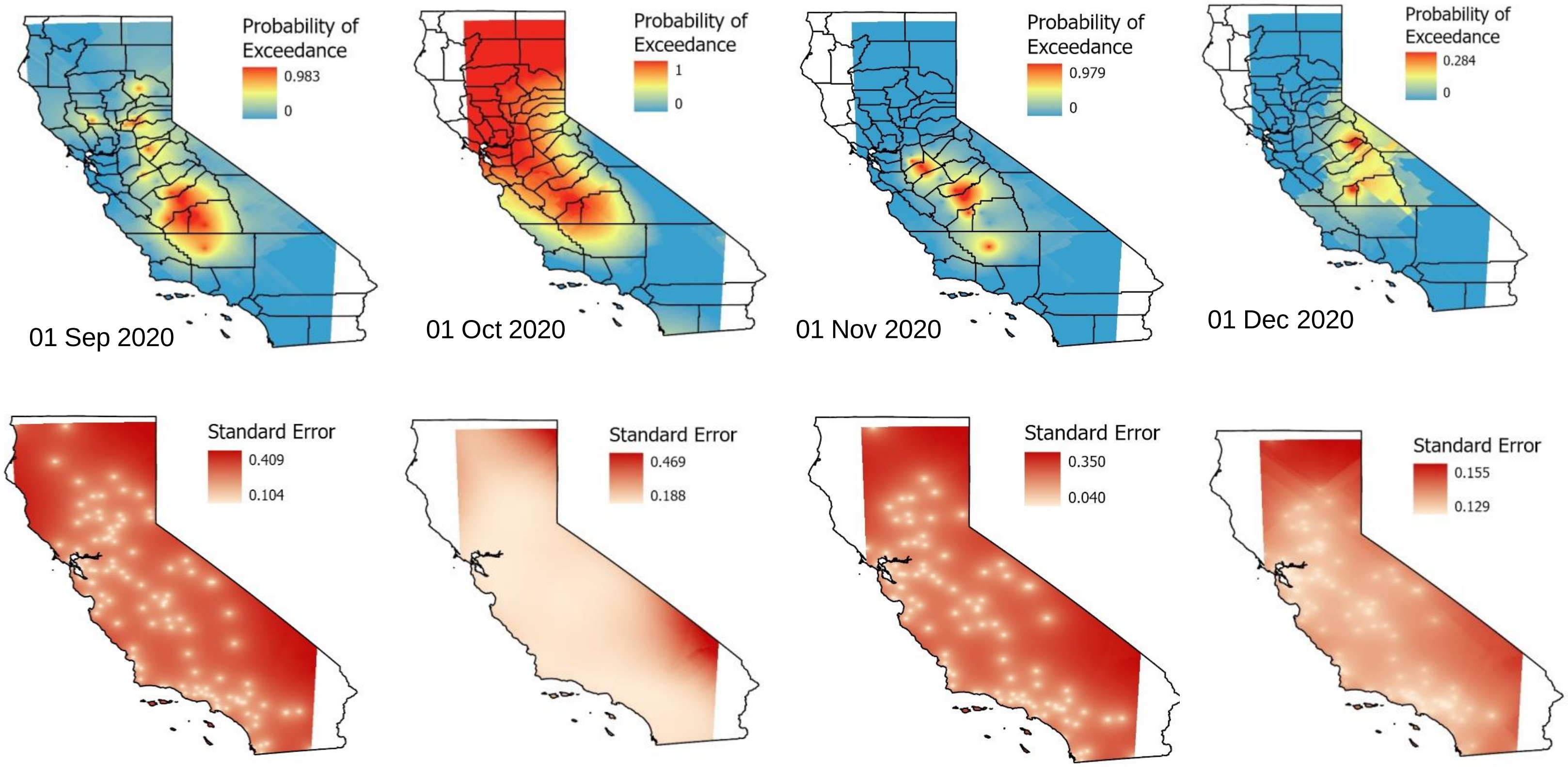
Pop_i represents the vulnerable population size for a given county i,

Pop_{min} represents the minimum vulnerable population for across all counties,

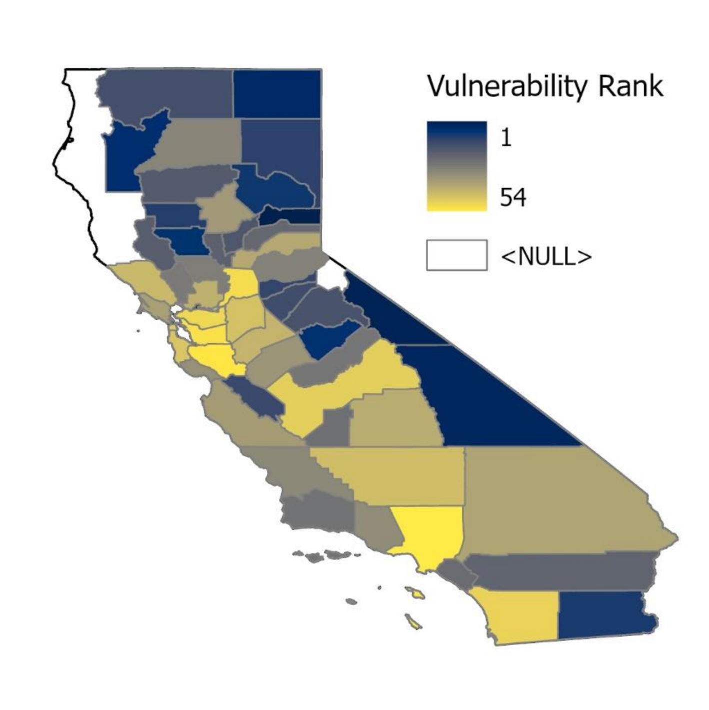
Pop_{max} represents the maximum vulnerable population for across all counties.



Results – IK Prediction/ SE Maps (P > 35.4 μg/m³)



Results – Vulnerability map (01 Oct 2020)



Higher Rank = More Vulnerable

Top 10 Most Vulnerable Californian Counties for 01 Oct 2020

County Name	Vul Pop Size	Oct01 IK Prob	Oct Vul Value	Oct01 Vul Rank
Los Angeles	4,268,959	0.2414	0.2414	54
Santa Clara	808,449	0.9630	0.1823	53
Alameda	692,333	1.0000	0.1621	52
Sacramento	638,122	1.0000	0.1494	51
Contra Costa	511,423	1.0000	0.1197	50
San Diego	1,371,935	0.2903	0.0933	49
Fresno	389,026	1.0000	0.0910	48
San Francisco	368,937	0.9487	0.0819	47
San Mateo	355,862	0.9590	0.0798	46
San Joaquin	296,544	1.0000	0.0693	45

Discussion

- Novel and complementary approaches - EBK and space-time cube (EHSA) to analyse and visualise PM2.5 data.
- From our analysis, we deduced that Sep-Oct are the periods with exceptionally high PM2.5 concentrations.
- EHSA – for each location how the hot/coldspot change with time – validates our interpolation results.
- Possible Reasons: 21 wildfires in Sep (mostly late Sep) & 16 wildfires in Oct; Valley temperature inversion hypothesis.
- Vulnerability maps – rank vulnerability of California counties to inform decision-making processes.
- Link to SDH concepts: uncertainty/ limitations.

Limitations: Uncertainty

Instrumentation and measurement uncertainty

Arises due to:
- Instrument breakdown

Solution:
- (Data collection) more frequent calibration and check up
- (Data user) perform data cleansing prior analysis

Interpolation Uncertainty

Arises due to:
- Nature of the data points (non-normality/ stationary)

Solution:
- Perform cross-validation to obtain RMSE
- Choose to use interpolation methods that can tolerate some abnormality in distribution

Contextual/ neighbourhood uncertainty

Arises due to:
- Emerging hotspot neighbourhood parameters

Solution:
- Domain expertise
- Perform spatial autocorrelation
- Not applicable for our case (fixed point sampling)

Conclusion

- Atmospheric pollutant studies usually use a variety of methods (no fixed methodology). We showcased a combination of complementary GIS methods to model, analyse, visualise and handle the uncertainty of spatially continuous datasets/ phenomenon.

- Achieved goals of mapping PM2.5 distribution, emerging hotspots, identified vulnerable groups, and quantified the corresponding uncertainties/ limitations of our study.

THANKS!

Questions?

References

Aguilera, R., Corringham, T., Gershunov, A., & Benmarhnia, T. (2021). Wildfire smoke impacts respiratory health more than fine particles from other sources: Observational evidence from Southern California. *Nature Communications*, *12*(1), 1493. https://doi.org/10.1038/s41467-021-21708-0

Aguilera, R., Gershunov, A., Ilango, S. D., Guzman-Morales, J., & Benmarhnia, T. (2020). Santa Ana Winds of Southern California Impact PM2.5 With and Without Smoke From Wildfires. *GeoHealth*, *4*(1), e2019GH000225. https://doi.org/10.1029/2019GH000225

Bashir, M. F., Ma, B. J., Bilal, Komal, B., Bashir, M. A., Farooq, T. H., Iqbal, N., & Bashir, M. (2020). Correlation between environmental pollution indicators and COVID-19 pandemic: A brief study in Californian context. *Environmental Research*, *187*, 109652. https://doi.org/10.1016/j.envres.2020.109652

Bhunia, G. S., & Ding, D. (2020). Temporal and spatial statistical analysis of ambient air quality of Assam (India). *Journal of the Air & Waste Management Association*, *70*(8), 775–794. https://doi.org/10.1080/10962247.2020.1772406

Brus, J., Voženílek, V., & Popelka, S. (2013, June 1). *An Assessment of Quantitative Uncertainty Visualization Methods for Interpolated Meteorological Data*. https://doi.org/10.1007/978-3-642-39649-6_12

Cisneros, R., Schweizer, D., Preisler, H., Bennett, D. H., Shaw, G., & Bytnerowicz, A. (2014). Spatial and seasonal patterns of particulate matter less than 2.5 microns in the Sierra Nevada Mountains, California. *Atmospheric Pollution Research*, *5*(4), 581–590. https://doi.org/10.5094/APR.2014.067

de Foy, B., Schauer, J. J., Helmig, D., & Goldstein, A. (2019). Changes in speciated PM2.5 concentrations in fresno, california, due to NOx reductions and variations in diurnal emission profiles by day of week. Elementa (Washington, D.C.), 7(1)https://doi.org/10.1525/elementa.384

Esri. (2021a). Emerging Hot Spot Analysis (Space Time Pattern Mining). Retrieved from: https://pro.arcgis.com/en/pro-app/latest/tool-reference/space-time-pattern-mining/emerginghotspots.htm

References

Esri. (2021b). Visualize the space-time cube. Retrieved from: https://pro.arcgis.com/en/pro-app/latest/tool-reference/space-time-pattern-mining/visualizing-cube-data.htm

Esri. (2021c). What is empirical Bayesian kriging? Retrieved from: https://pro.arcgis.com/en/pro-app/latest/help/analysis/geostatistical-analyst/what-is-empirical-bayesian-kriging-.htm

Fang, T. B., & Lu, Y. (2011). Constructing a Near Real-time Space-time Cube to Depict Urban Ambient Air Pollution Scenario. *Transactions in GIS*, *15*(5), 635–649. https://doi.org/10.1111/j.1467-9671.2011.01283.x

Huang, F., Li, X., Wang, C., Xu, Q., Wang, W., Luo, Y., Tao, L., Gao, Q., Guo, J., Chen, S., Cao, K., Liu, L., Gao, N., Liu, X., Yang, K., Yan, A., & Guo, X. (2015). PM2.5 Spatiotemporal Variations and the Relationship with Meteorological Factors during 2013-2014 in Beijing, China. *PLOS ONE*, *10*(11), e0141642. https://doi.org/10.1371/journal.pone.0141642

Keler, A., & Krisp, J. M. (2015). Spatio-temporal Visualization of Interpolated Particulate Matter (PM2.5) in Beijing. *GI_Forum*, *1*, 464–474. https://doi.org/10.1553/giscience2015s464

Li, L., Girguis, M., Lurmann, F., Pavlovic, N., McClure, C., Franklin, M., Wu, J., Oman, L. D., Breton, C., Gilliland, F., &Habre, R. (2020). Ensemble-based deep learning for estimating PM2.5 over California with multisource big data including wildfire smoke. *Environment International*, *145*, 106143. https://doi.org/10.1016/j.envint.2020.106143

Luo, J., Du, P., Samat, A., Xia, J., Che, M., & Xue, Z. (2017). Spatiotemporal Pattern of PM2.5 Concentrations in Mainland China and Analysis of Its Influencing Factors using Geographically Weighted Regression. *Scientific Reports*, *7*(1), 40607. https://doi.org/10.1038/srep40607

References

Rooney, B., Wang, Y., Jiang, J. H., Zhao, B., Zeng, Z.-C., & Seinfeld, J. H. (2020). Air quality impact of the Northern California Camp Fire of November 2018. *Atmospheric Chemistry and Physics*, *20*(23), 14597–14616. https://doi.org/10.5194/acp-20-14597-2020

Senaratne, H., Gerharz, L., Pebesma, E., & Schwering, A. (2012). Usability of Spatio-Temporal Uncertainty Visualisation Methods. In J. Gensel, D. Josselin, & D. Vandenbroucke (Eds.), *Bridging the Geographic Information Sciences* (pp. 3–23). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-29063-3_1

Stowell, J. D., Bi, J., Al-Hamdan, M. Z., Lee, H. J., Lee, S.-M., Freedman, F., Kinney, P. L., & Liu, Y. (2020). Estimating $PM_{2.5}$ in Southern California using satellite data: Factors that affect model performance. *Environmental Research Letters*, *15*(9), 094004. https://doi.org/10.1088/1748-9326/ab9334

The United States Environmental Protection Agency (U.S. EPA). (2012). Revised air quality standards for particle pollution and updates to the Air Quality Index (AQI) Retrieved from: https://www.epa.gov/sites/production/files/201604/documents/2012_aqi_factsheet.pdf

The United States Environmental Protection Agency (U.S. EPA). (2021). Retrieved from:https://www.epa.gov/expobox/exposure-assessment-tools-lifestages-and-populationshighly-exposed-or-other-susceptible

Vodonos, A., & Schwartz, J. (2021). Estimation of excess mortality due to long-term exposure to PM2.5 in continental United States using a high-spatiotemporal resolution model. *Environmental Research*, *196*, 110904. https://doi.org/10.1016/j.envres.2021.110904

Wang, T., Zhao, B., Liou, K.-N., Gu, Y., Jiang, Z., Song, K., Su, H., Jerrett, M., & Zhu, Y. (2019). Mortality burdens in California due to air pollution attributable to local and nonlocal emissions. *Environment International*, *133*, 105232. https://doi.org/10.1016/j.envint.2019.105232

World Health Organisation (WHO). (2004). Health aspects of air pollution: results from the WHO project" Systematic review of health aspects of air pollution in Europe".

References

Xiao, F., Yang, M., Fan, H., Fan, G., & Al-qaness, M. A. A. (2020). An improved deep learning model for predicting daily PM2.5 concentration. *Scientific Reports*, *10*(1), 20988. https://doi.org/10.1038/s41598-020-77757-w

Yang, Q., Yuan, Q., Li, T., Shen, H., & Zhang, L. (2017). The Relationships between PM2.5 and Meteorological Factors in China: Seasonal and Regional Variations. *International Journal of Environmental Research and Public Health*, *14*(12), 1510. https://doi.org/10.3390/ijerph14121510

Zhang, H., Wang, Z., & Zhang, W. (2016). Exploring spatiotemporal patterns of PM2.5 in China based on ground-level observations for 190 cities. *Environmental Pollution*, *216*, 559–567. https://doi.org/10.1016/j.envpol.2016.06.009