Estimation of ambient $PM_{2.5}$ in Iraq and Kuwait from 2001 to 2018 using machine learning and remote sensing

Jing Li

Environment International 2021

Introduction

- Since 2001, nearly 3 million U.S. military personnel and coalition military personnel have been deployed in support of operations in the Middle East with frequent exposure to high PM.

- Although land-based military personnel deployed to Iraq and Kuwait were exposed to high levels of particulate matter (PM), the lack of a ground-based PM monitoring network has limited the assessment of health.

- **The aim of this study is assessing the daily $PM_{2.5}$ exposures for Iraq and Kuwait.**

- Satellite-based AOD has been widely used to estimate spatial-temporarily resolved $PM_{2.5}$ exposure based on ground-level $PM_{2.5}$ measurements.

- **However, this method has not been applied to areas with a paucity of $PM_{2.5}$ monitoring sites.**

- We developed a calibration model to convert visibility to $PM_{2.5}$ based on satellite AOD. This approach takes advantage of the large database of historical visibility collected by airport and the high spatial-temporal resolution of satellite-based AOD.

Data Collection : 2001-2018

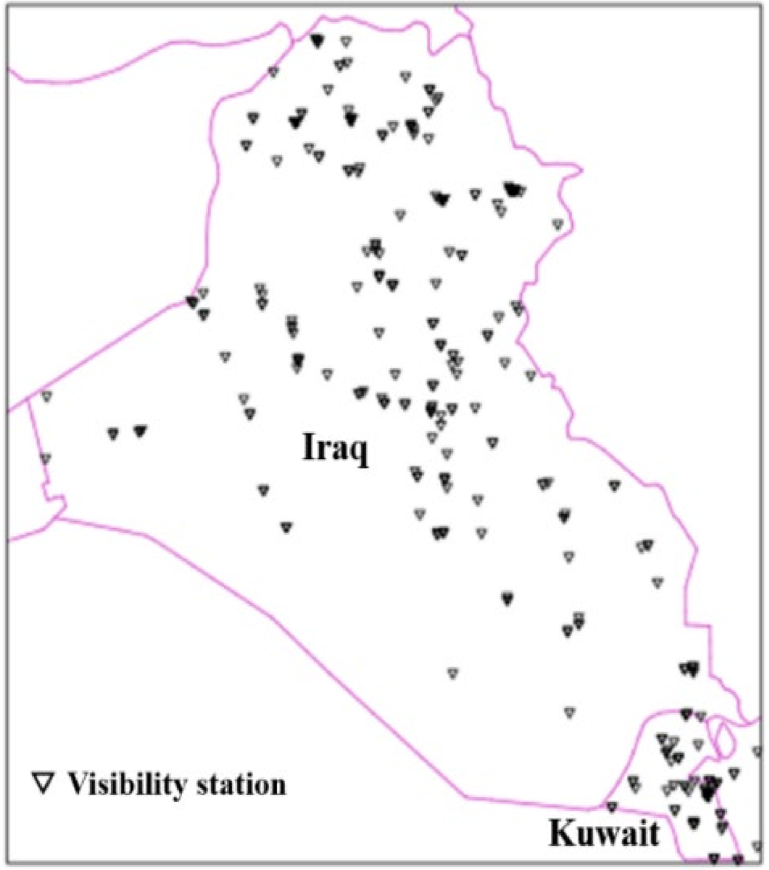
Data		Data source	Spatial resolution	Temporal resolution
AOD		NASA MAIAC	1km	Daily
Visibility		United States air force airport	780 sites	Hourly
$PM_{2.5}$	Kuwait: 2004-2005, 2017-now	Harvard	3 sites for 2004-2005; 2 site for 2017-2019	Around 2000 samples
	Kuwait: 2017-2018	EPA U.S. Embassy Kuwait	1 site	
	Iran	Iran cooperator	90 sites	Daily
Dust	Dust Surface Mass Concentration	NASA MERRA-2	0.625°×0.5°	Daily
	Dust Extinction AOD			
	Dust Scattering AOD			
	Dust Column Mass Density			
Land use	NDVI	NOAA AVHRR	5 km	Daily
	Road density	OpenStreetMap	1 km	--
	Distance to industrial area	U.S. Geospatial-Intelligence Agency	1 km	--
	Elevation	NOAA	1 arc-minute	--

Data Collection : 2001-2018

Data		Data source	Spatial resolution	Temporal resolution
Meteorological data	Temperature at 2 m	ERA-5 Reanalysis produced by European Centre for Medium-Range Weather Forecasts	31 km	Daily
	U-wind speed at 10 m			
	V-wind speed at 10 m			
	Instantaneous 10m wind gust			
	Dew point temperature at 2m			
	Total precipitation			
	Surface pressure			
	Downward UV radiation			
	Planetary boundary layer height			
	Total cloud cover			
	Low cloud cover			
	Medium cloud cover			
	High cloud cover			
	High vegetation cover			
	Low vegetation cover			
	Forecast albedo			
	Evaporation			
	Relative humidity			

Data Collection : 2001-2018

Map of visibility stations

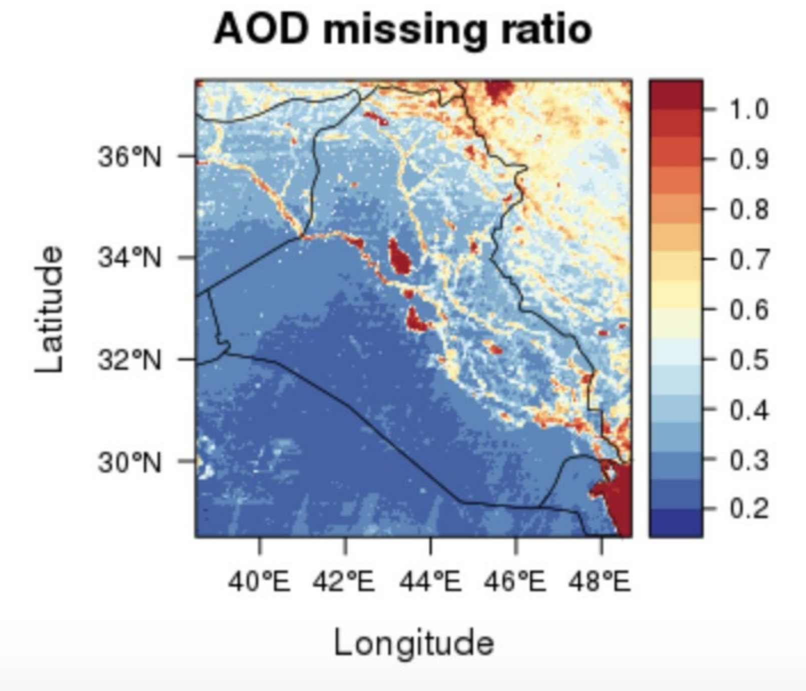


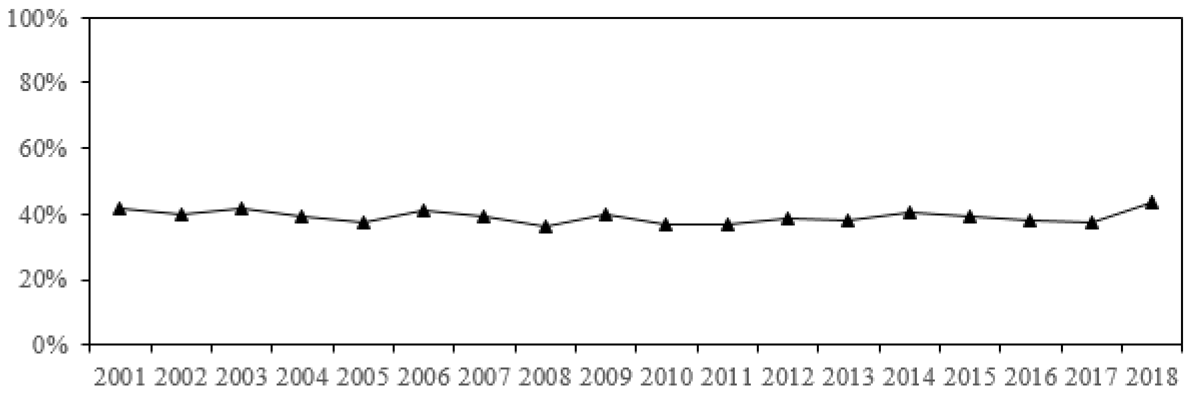
$PM_{2.5}$ sampling information in Kuwait.

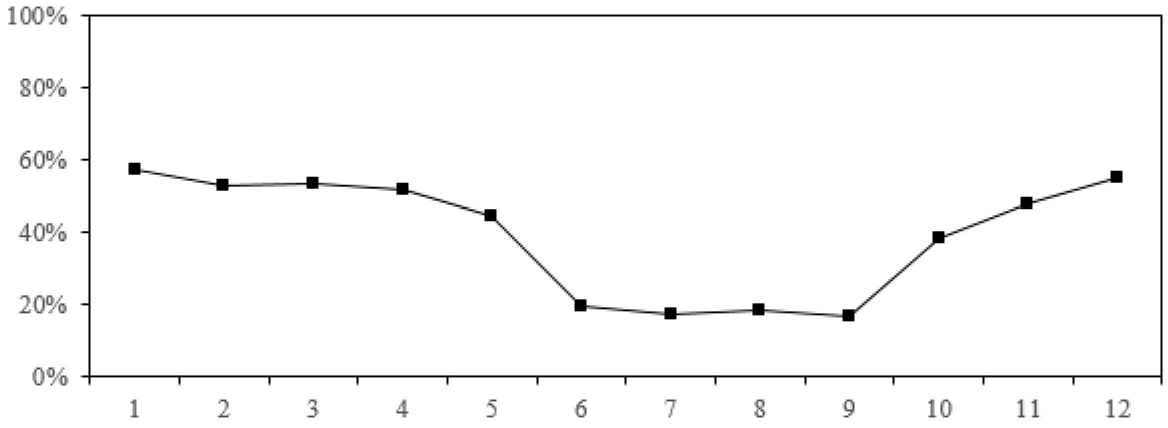
Sampling site	Longitude, Latitude	Sampling date	Number of Sample
Central site	29.33,47.97	2004-2005	439
		2017-2019	395
Northern site	29.77,47.77	2004-2005	40
Southern site	28.96,48.16	2004-2005	62
		2017-2019	350
U.S. Embassy Kuwait	29.31,48.04	2017-2018	656

A total of 1942 valid $PM_{2.5}$ observations were used in this study.

Data Collection : 2001-2018







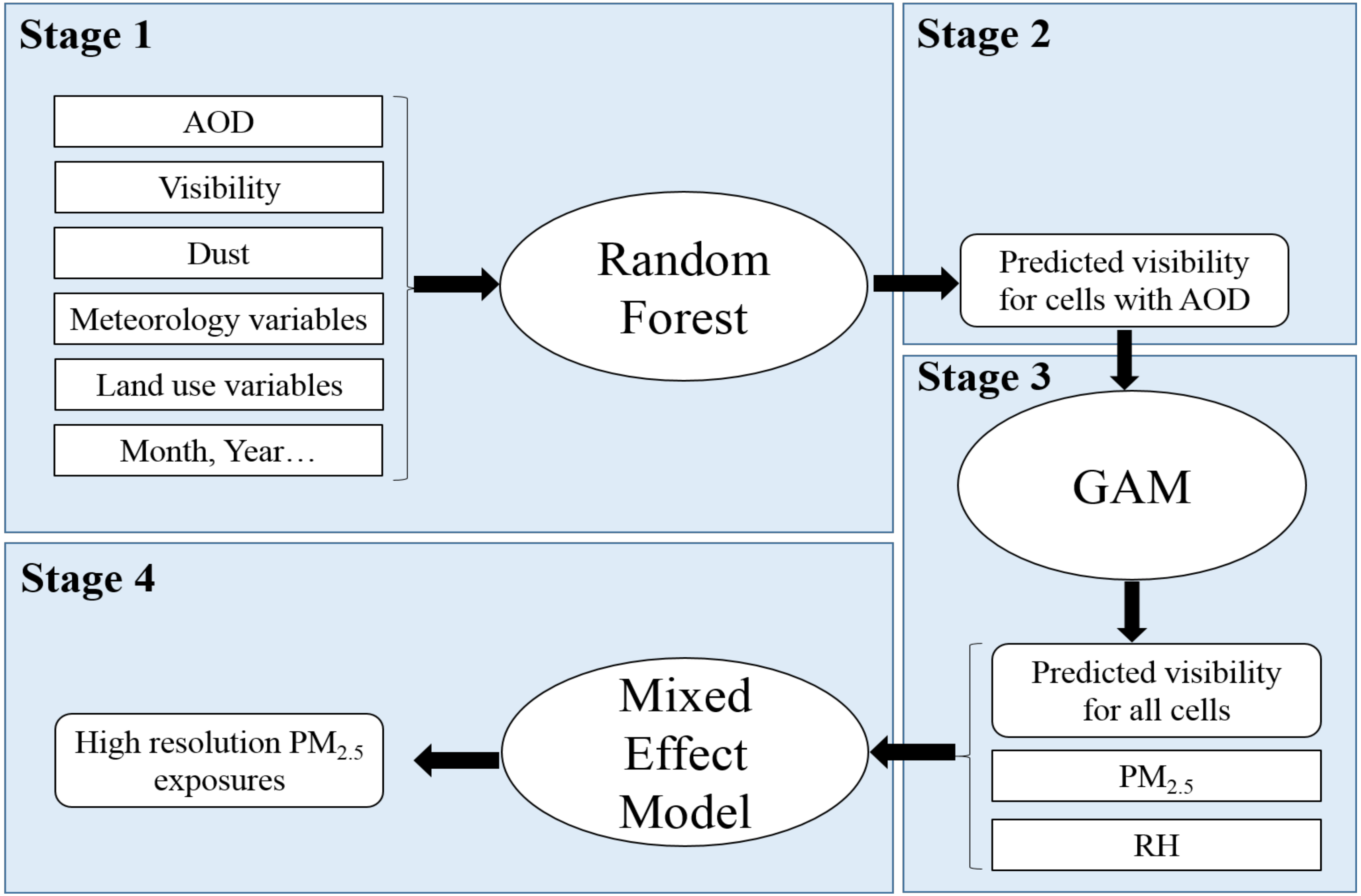
The percentage of the missing AOD values in each grid cell from 2001-2018.

The percentage of the missing AOD values by month and by year.

Model

In order to estimate daily $PM_{2.5}$ concentrations in each $1km^2$ grid cell, we developed a hybrid model including four stages.

Flowchart of the four-stage modeling approach.



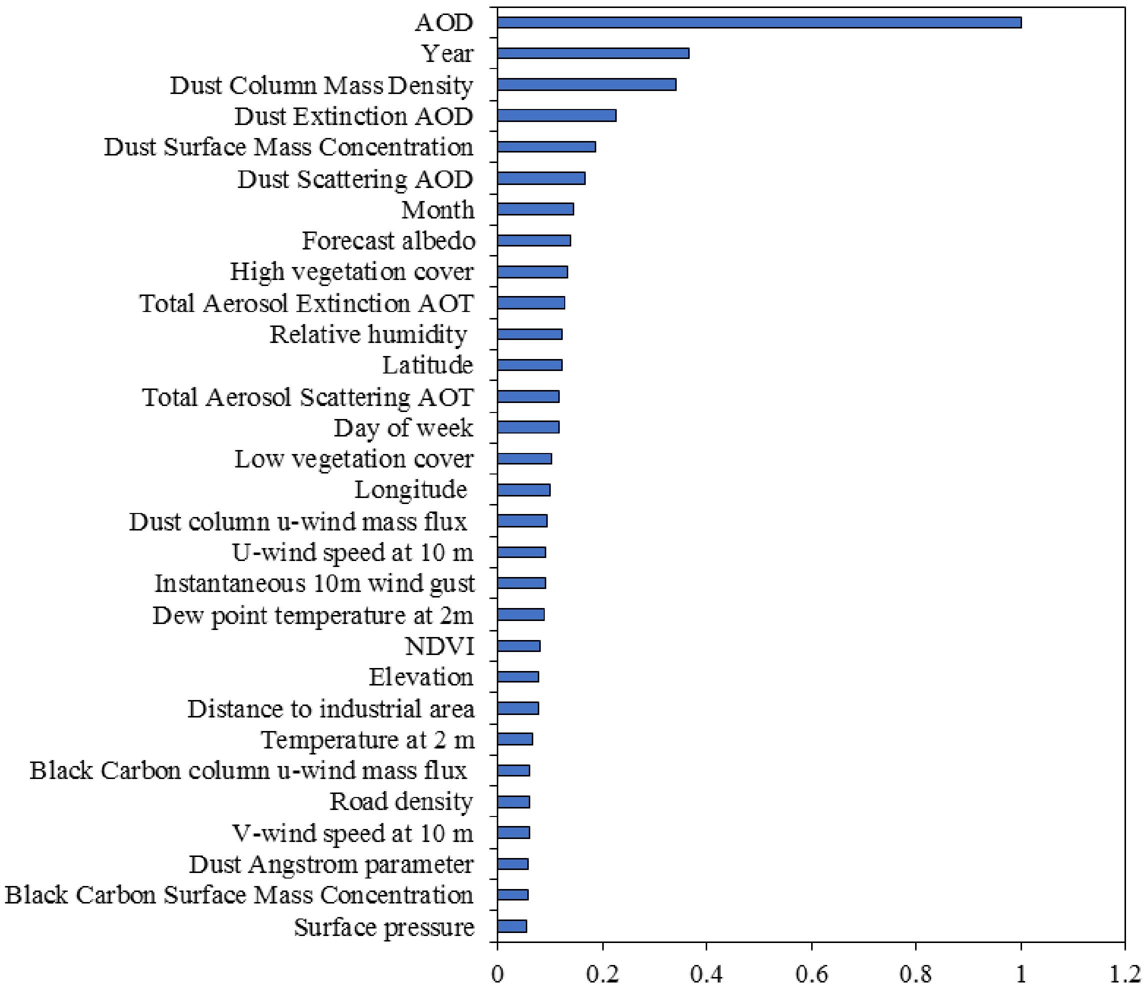
Model performance

Stage 1: Random forest model VIS~AOD

We performed "out of sample" 10-fold cross-validation to estimate the visibility prediction model results and avoid over fitting. All visibility stations were randomly divided into 10%-90% splits.

The CV R^2 between fitted and predicted daily visibility was 0.71

MAIAC AOD, year, dust column mass density, dust extinction AOD, and dust surface mass concentration are the five most important predictors.



Top 30 most importance variables for the random forest model predicting daily visibility.

Model performance

Stage 2: For grid cells with AOD but without visibility measurements, we used the Stage 1 model to estimate daily visibility for each grid.

Stage 3: For grid cells missing both AOD and visibility measurements, a generalized additive mixed model (GAMM) was used to predict daily visibility based on the output from Stage 2 model

$$PredVis_{ij} = (\alpha + \mu_i) + (\beta + v_i)MVis_j + Smooth(X, Y)_{k(j)} + \epsilon_{i,j}$$

$$(u_j v_j) \sim \left((00), \Omega_\beta\right) \qquad \text{Eq. (1)}$$

The stage 3 model performed well with a mean out of sample R^2 of 0.68.

Model performance

Stage 4: Mixed effect model $PM_{2.5}$~VIS

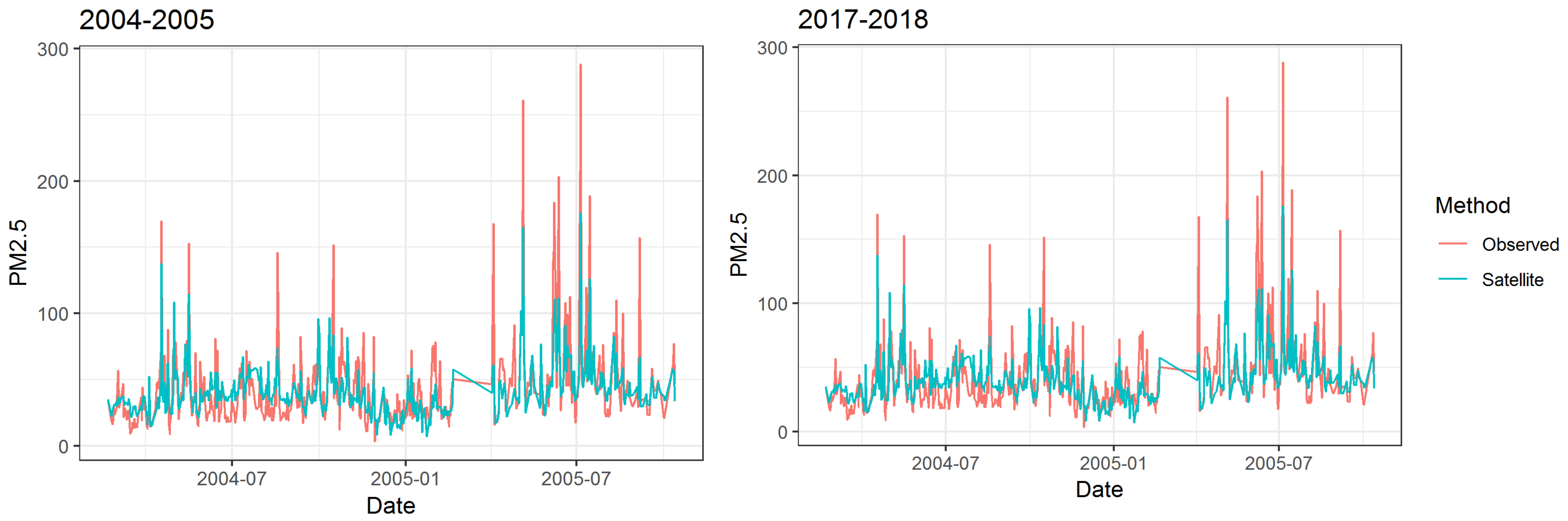
We used data during 2017-2018 to train the model and use data during 2004-2005 for evaluation, which can help to understand the model's ability to estimate historical $PM_{2.5}$.

Model variables	Cross validation R^2 in the modeling year (2017-2018)	Validation R^2 of historical estimates at daily levels (2004-2005)
VIS	0.600	0.635
1/VIS	0.664	0.723
lnVIS	0.656	0.698
RH+1/VIS	0.662	0.724
WS+1/VIS	0.667	0.724
TEMP+1/VIS	0.630	0.685
RH+TEMP+1/VIS	0.669	0.722
RH+WS+1/VIS	0.637	0.705
TEMP+WS+1/VIS	0.673	0.721
RH+TEMP+WS+1/VIS	0.632	0.686
RH+RH2+TEMP+WS+1/VIS	0.706	0.743
RH+RH^2+1/VIS	0.698	0.745

This model resulted in a high CV R^2 value of 0.70 in the modeling year (2017-2018) and R^2 value of 0.74 in the evaluating year (2004-2005),indicating good model

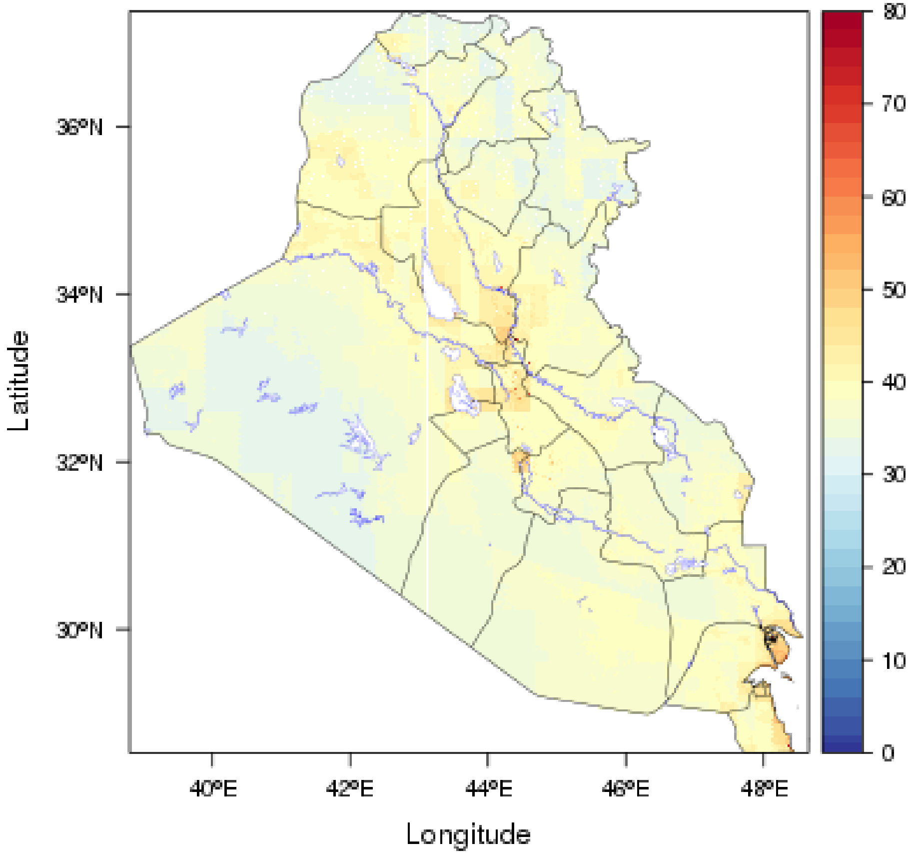
Model performance

Long-term variation of satellite-derived and ground-observed $PM_{2.5}$.



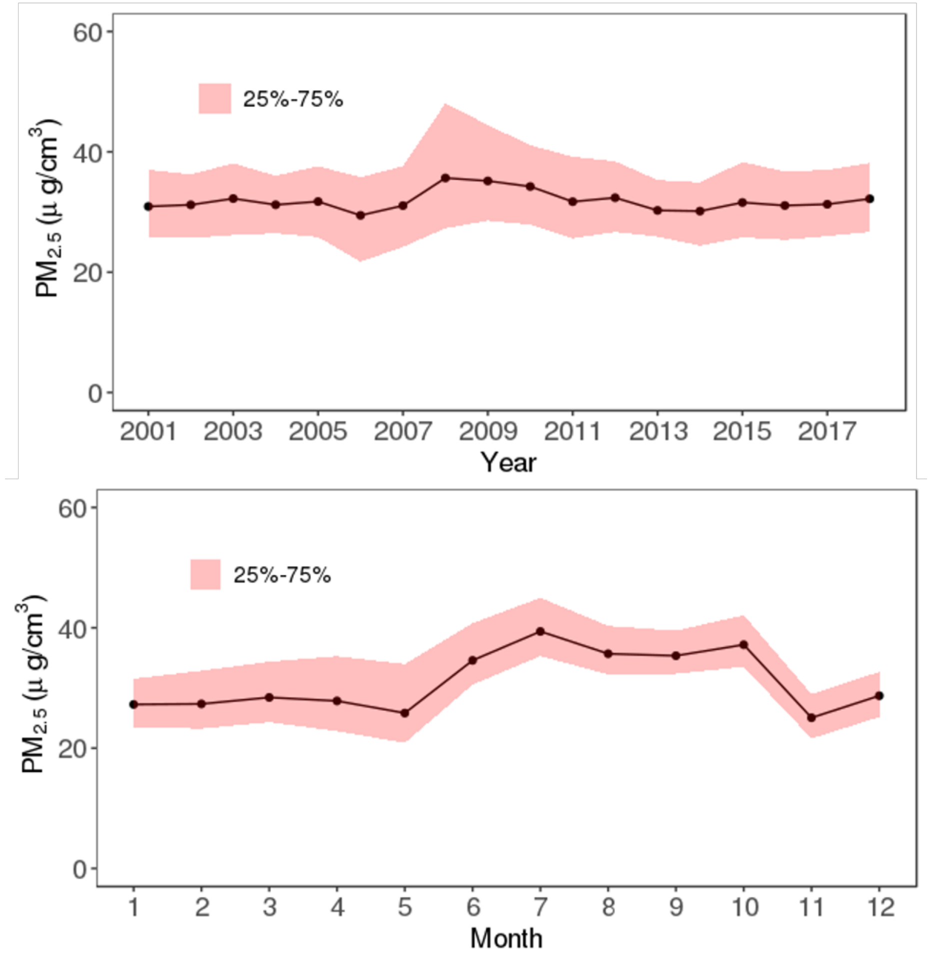
Results

Mean $PM_{2.5}$ concentrations ($\mu g/m^3$) in each 1 km × 1km grid during the entire modeling



Some big cities such as Bagdad, Karbala, and Najaf, and Diwaniya have annual average $PM_{2.5}$ concentrations above 45 ug/m^3, areas of Iraq that included many U.S. bases. In contrast, Al-Anbar, the largest governorate west of Iraq which is mostly desert showed low average $PM_{2.5}$ concentrations, except for large urban areas in the east (Al-Fallujah and Al-Ramadi). Some bases were located in regions with lower $PM_{2.5}$ concentrations such as in the Syrian Desert, which have average $PM_{2.5}$ concentrations below 30 ug/m^3.

Results

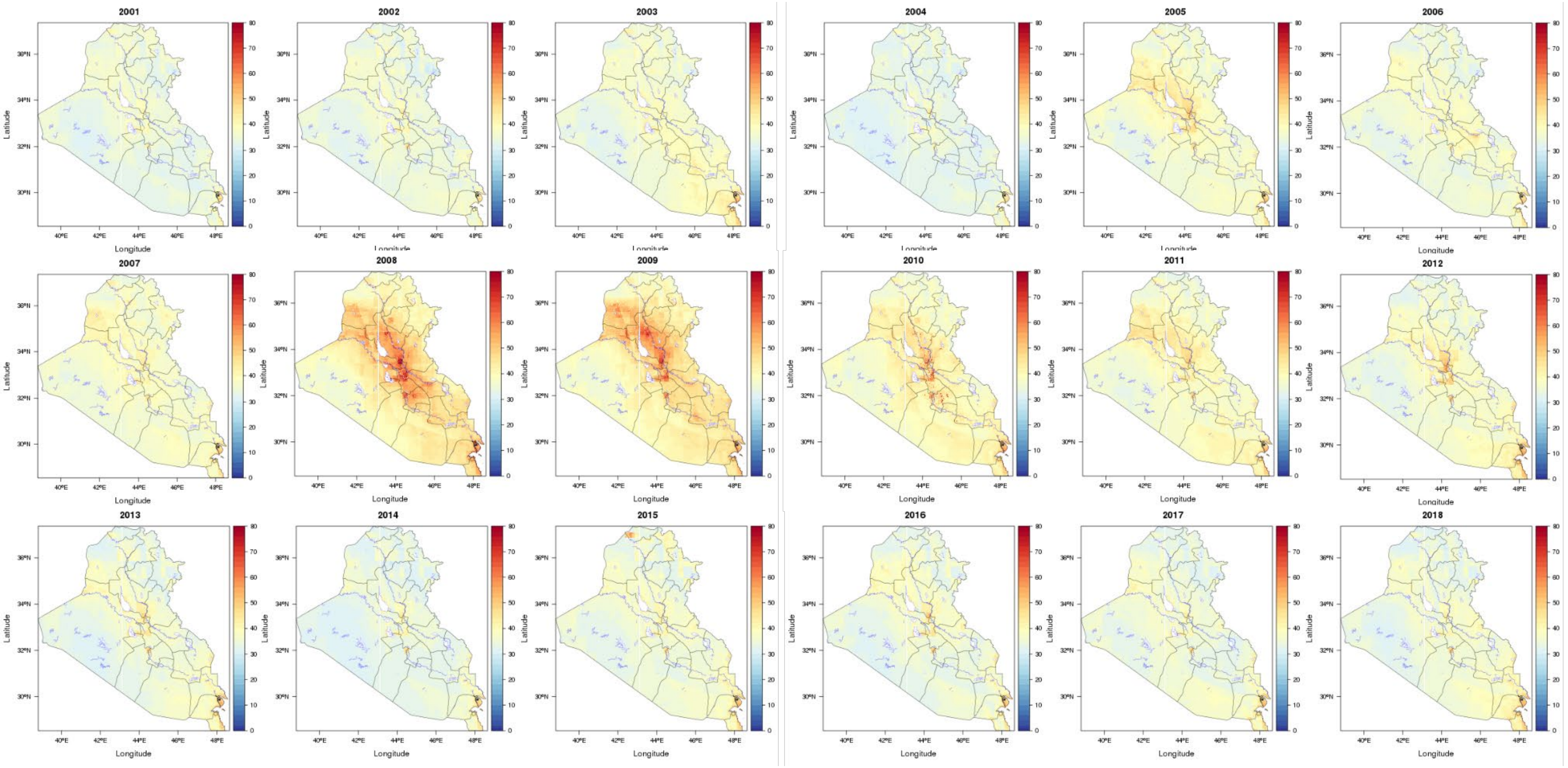


- Mean values for Iraq and Kuwait were 36.85 and 40.78 μg/m^3, respectively.
- The highest annual $PM_{2.5}$ concentration for both Kuwait and Iraq were observed in 2008, followed by 2009. It is possible that the extreme drought in 2008-2009 contributed to the high $PM_{2.5}$ level in 2008.
- July had the highest predicted concentrations, and November had the lowest values.

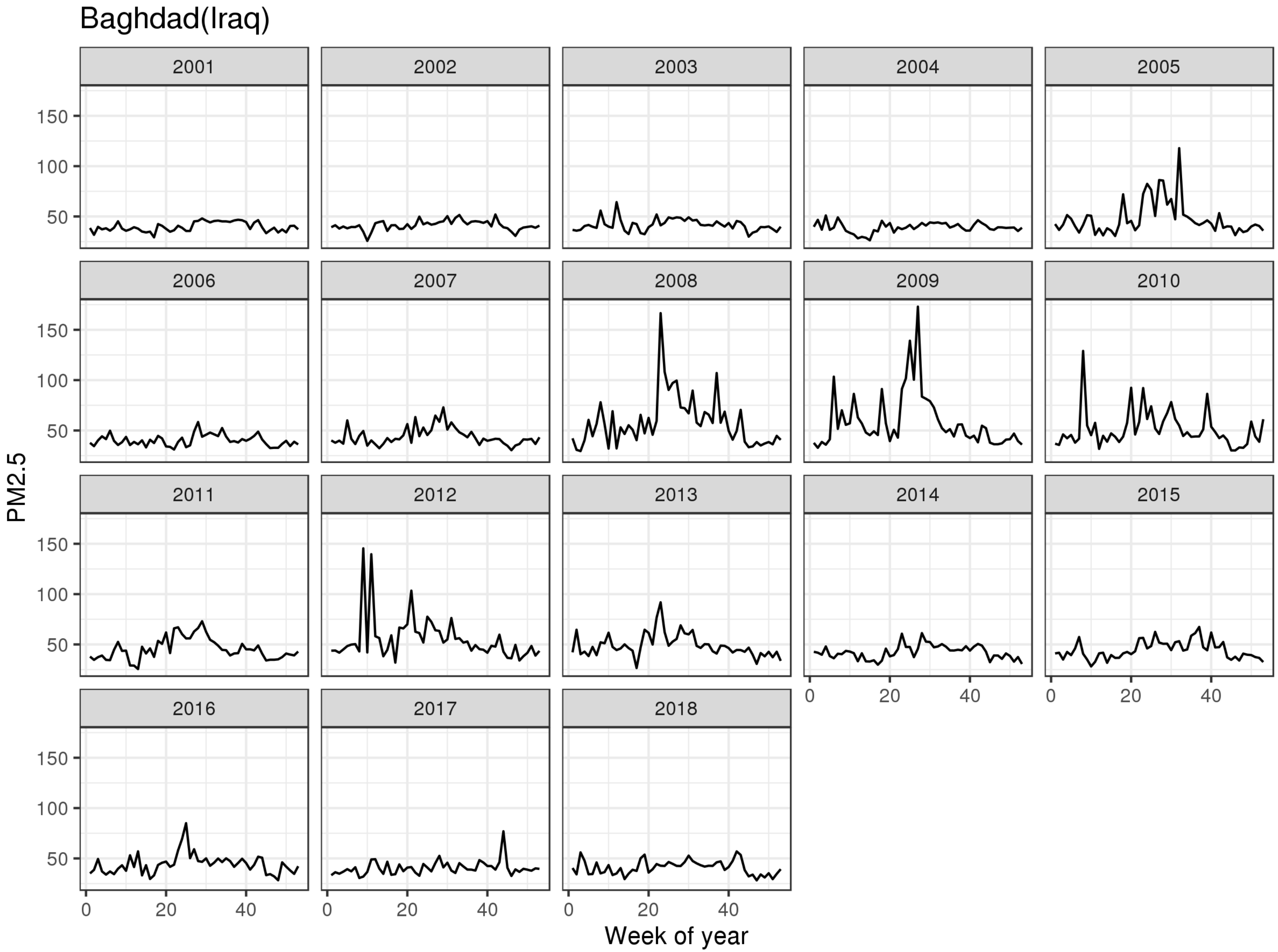
Annual and monthly median $PM_{2.5}$ concentrations in Kuwait and Iraq.

Result

Spatial distribution of predicted mean $PM_{2.5}$ concentrations (μg/m^3) for each year

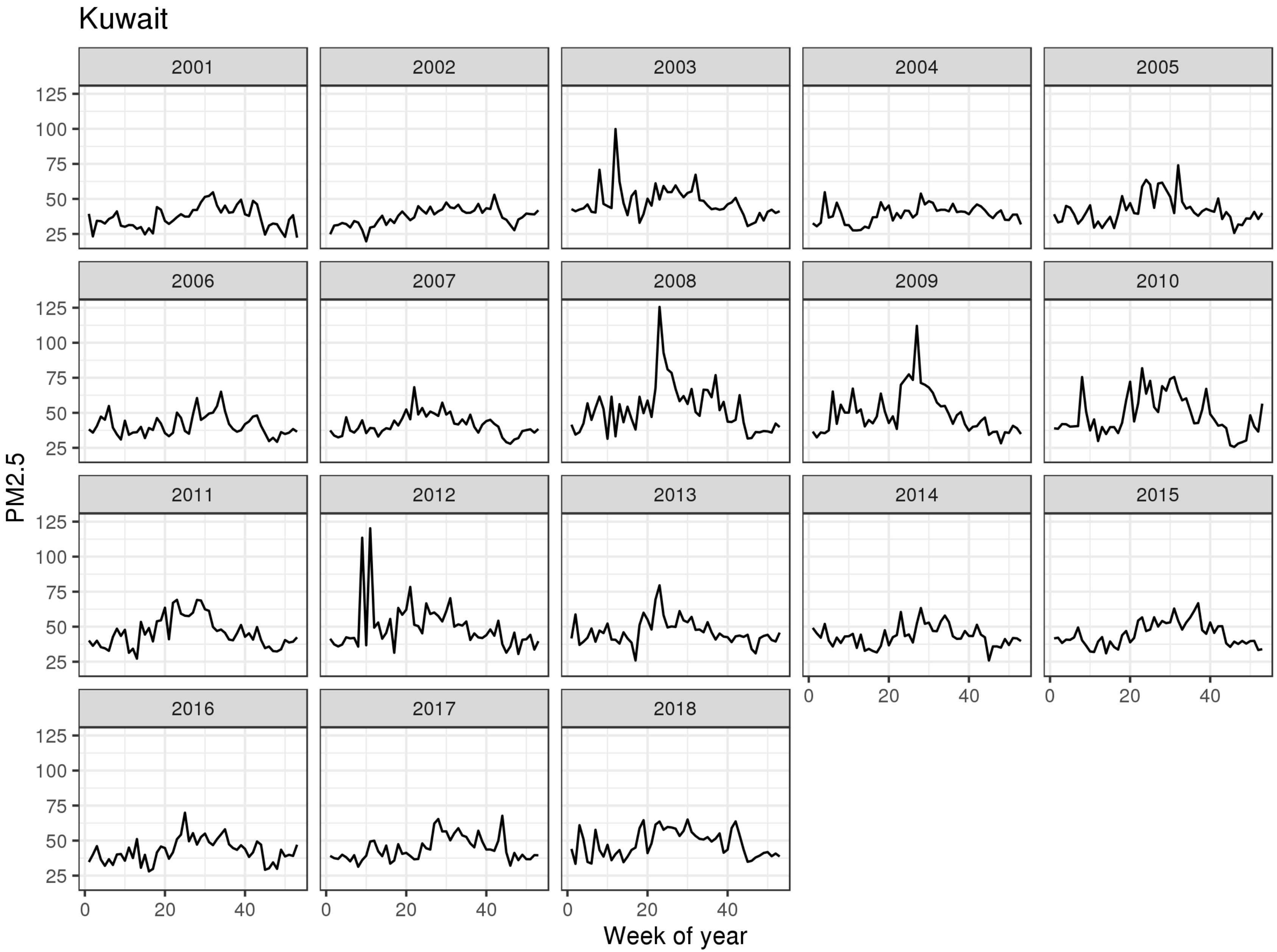


Result



Weekly average $PM_{2.5}$ concentrations at Baghdad International Airport.

Result



Weekly average $PM_{2.5}$ concentrations at Kuwait International Airport

Discussion

Comparisons

- The annual mean predicted $PM_{2.5}$ concentrations for Kuwait (40.78μg/m^3) and Iraq (36.85 μg/m^3) from 2001-2018 **were much higher than the U.S. National Air Quality Standard** of 12 μg/m^3 and the WHO guideline of 10 μg/m^3.
- Very few studies conducted PM measurements in Kuwait or Iraq. Al-Hemoud et al. (2018a) shown daily air concentration of PM_{10} for the year 2012 in a suburban area of Kuwait. Al-Hemoud et al. (2018b) and Al-Hemoud et al. (2019) collected $PM_{2.5}$ measurements in three monitoring stations in Kuwait from 2014-2017. Our predictions and the PM_{10} in Al-Hemoud et al. (2018a) shown the same seasonal pattern. The $PM_{2.5}$ concentrations predicted in this study are consistent with previous WHO air pollution database, but the $PM_{2.5}$ concentrations in Al-Hemoud et al. (2019) are higher than our results. It is possible because only the hourly $PM_{2.5}$ values above the minimum background concentration of 8.83 μg/m^3 were chosen in their study. The seasonal trends of the $PM_{2.5}$ are consistent with the dust storms in the region, which may impact the monthly means of $PM_{2.5}$ concentrations.

Limitations

- A limitation of our approach is that the stage 4 model regression results was based on $PM_{2.5}$ measurement data **collected at three sampling sites for two years** because of the paucity of $PM_{2.5}$ data in this region.
- Another limitation is that the R^2 of cross-validation for our daily $PM_{2.5}$ prediction is not as high as those from models for other countries such as United States (R^2 = 0.84) **due to the lack of key data sources** (emission inventories, CTM outputs).

Discussion

Advantages

Our study has some major findings that are of public health significance.

- In previous studies, the AOD-$PM_{2.5}$ relationship was directly used to examine the spatial distribution of $PM_{2.5}$. However, for many countries, there are no available historical $PM_{2.5}$ measurements. Our novel exposure assessment approach **indicates it possible to assess air pollution exposures in countries without extensive $PM_{2.5}$ monitoring locations**, and without extensive $PM_{2.5}$ historical data. And the other predictors used in our model are predominantly from global public datasets, thus our method could be applied in other countries.
- Secondly, the random forest model approach **provided the relative importance of each factor associated with historical $PM_{2.5}$ concentrations** to assess exposures in this region or similar areas. For example, in addition to AOD, we found that surface dust related variables are also associated with $PM_{2.5}$ in arid environments (Li et al., 2020).

- To the best of our knowledge, **this is the first model** to estimate historical $PM_{2.5}$ concentrations at a high resolution in this region.

Thank you