Merging MODIS and Ground-Based Fine Mode Fraction of Aerosols Based on the Geostatistical Data Fusion Method

Aimei Zhao [1,2], Zhengqiang Li [1,*], Ying Zhang [1,*] [iD], Yang Zhang [1,2] [iD] and Donghui Li [1]

[1] Environment Protection Key Laboratory of Satellite Remote Sensing, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100101, China; zhaoam@radi.ac.cn (A.Z.); zhangyang2015@radi.ac.cn (Y.Z.); lidh@radi.ac.cn (D.L.)

[2] University of Chinese Academy of Sciences, Beijing 100049, China

* Correspondence: lizq@radi.ac.cn (Z.L.); zhangying02@radi.ac.cn (Y.Z.)

Abstract: With the rapid development of the economy and society, fine particulate matter ($PM_{2.5}$) has not only caused severe environmental problems, but also posed a threat to public health. In order to improve the estimated accuracy of $PM_{2.5}$, the input data fine mode fraction (FMF), a key parameter to the $PM_{2.5}$ remote sensing method (PMRS), should be improved due to its significant errors. In this study, we merge the observations of the fine mode fraction (FMF) from the Moderate Resolution Imaging Spectroradiometer (MODIS), the Aerosol Robotic Network (AERONET) and the Sun-sky radiometer Observation Network (SONET) using the universal kriging (UK) method to obtain accurate FMF distribution over eastern China. $PM_{2.5}$ mass concentration is estimated by the fusion and MODIS FMF distributions using the PMRS model. The results show that the parameters in the variogram are relatively stable except for significant differences in correlation lengths in summer. The FMF in the Winter of 2015 shows that the mean error decreases from 0.38 to 0.13 compared with that from MODIS using leave-one-out cross-validation, with the maximum error decreasing from 0.75 to 0.34, indicating that the UK method can provide better estimates of FMF. We also find that $PM_{2.5}$ estimated from FMF fusion results is closer to the in situ $PM_{2.5}$ from the Ministry of Environmental Protection (MEP) (87.2 vs. 88.9 $\mu g/m^3$).

Keywords: FMF; UK fusion; $PM_{2.5}$; PMRS; MODIS

1. Introduction

In recent years, atmospheric particulate matter has reached a high level due to human activities. $PM_{2.5}$ (atmospheric particulate matter with a mass median diameter less than 2.5 μm), which has a great influence not only on air quality but also on human health, has drawn a wide range of attention [1–3]. In order to obtain the spatial distribution of $PM_{2.5}$ mass concentration near the ground, many studies have developed three kinds of methods based on remote sensing observations. These include the statistical method [4–9], the simulated method coupled with the atmospheric chemical model [10–12], and the physical dependent method [13]. The statistical method can be constrained by the model's validation since it needs a long data period and must be re-validated in different study areas. The simulated method is affected by many factors from atmospheric chemical models such as the simulation scale, chemical mechanisms, and emission inventories. The physical dependent method is a semi-physical model based on remote sensing observations to estimate $PM_{2.5}$ mass concentration near the ground, which effectively solves the limitations of the above two methods. In this physical dependent method, fine mode fraction (FMF, defined as the fine-mode fraction to the total aerosol optical depth (AOD)) is applied as a key parameter to distinguish the $PM_{2.5}$ from

the total suspended particles. However, the high uncertainty of FMF from satellite (mainly from the Moderate Resolution Imaging Spectroradiometer (MODIS) sensors) is a problem demanding prompt solution [10].

Data fusion approaches can effectively improve the accuracy of current FMF products. Many studies have attempted to merge the AOD from multi-sensors (e.g., MODIS, Multi-angle Imaging SpectroRadiometer (MISR), etc.) [14–16] with ground-based data (e.g., Aerosol Robotic Network (AERONET), etc.) [17–19] using the prevalent fusion approaches, including least squares [20], the maximum likelihood estimate method [21–24], the universal kriging (UK) method [25], and the spatial statistical data fusion (SSDF) method [26], etc. Compared with the approaches listed, the UK method is not only a linear and unbiased method, but also takes spatial correlation into account. Moreover, it is simple and easy to operate and has low computational complexity [27]. However, based on the UK method, the past studies mainly focus on merging AOD data from the multi-sensor or the multi-sensor with an AERONET product, with a lesser focus on FMF fusion.

In order to improve the accuracy of MODIS FMF over land, we study the feasibility of merging satellite data (MODIS) with ground-based data (AERONET and Sun-sky radiometer Observation Network (SONET)) to improve the accuracy of FMF. MODIS FMF and ground-based FMF observations of eastern China in the Winter (December, January, and February) of 2015 are merged to test the applicability based on the UK method. The FMF fusion results are then applied to the $PM_{2.5}$ remote sensing method (PMRS) to estimate $PM_{2.5}$ mass concentration near the ground. In Section 2, the methods and data are described in detail. The spatiotemporal variability analysis, FMF fusion test and cross-validation, and the estimated $PM_{2.5}$ results are introduced in Section 3. Section 4 gives the discussion, conclusion, and suggestions for possible improvements in the future.

2. Methods and Data

2.1. Comparison of Theretical FMFs

Currently, FMF can be obtained only by the MODIS product in space-borne remote sensing. In order to merge the FMF from satellite and ground-based products, the theoretical FMFs in different methods firstly were compared with each other. So far, FMF can be calculated by three retrieval methods, including the mode combination method, the truncation radius method, and the spectral identification method [28–32]. In the study, only two (mode combination and spectral identification) methods are used to calculate the FMFs. The FMF in the mode combination method used for MODIS is determined by different combinations of fine and coarse modes to match the observed spectral reflectance.

$$FMF_{MODIS} = \frac{\tau^{f}_{0.55}}{\tau^{tot}_{0.55}} \tag{1}$$

where $\tau^{f}_{0.55}$ is fine optical depth at 0.55 μm and $\tau^{tot}_{0.55}$ is the total aerosol optical depth at 0.55 μm. Additionally, the spectral identification method, also called the spectral deconvolution algorithm (SDA), calculates FMF based on the particle spectral characteristics. It is the standard algorithm of the AERONET FMF product. The SDA method is expressed as follows:

$$FMF_{SDA} = \frac{\alpha - \alpha_c}{\alpha_f - \alpha_c} \tag{2}$$

where α is the Ångström index and the subscripts f and c represent the fine and coarse modes, respectively. In order to capture the differences between the mode combination and the SDA method, we calculate the FMFs of three typical aerosol models, including the water-soluble (WS), biomass burning (BB), and dust (DU) models. The aerosol size distribution is:

$$\frac{\mathrm{dV}(r)}{\mathrm{dlnr}} = \sum_{i=f,c} \left[c_i \left((2\pi)^{\frac{1}{2}} |ln\sigma_i| \right) \right] \cdot \exp\left(-\frac{(lnr - lnR_i)^2}{2ln^2\sigma_i} \right) \tag{3}$$

where dV/dlnr (in units of $\mu m^3/\mu m^2$) is the aerosol volume particle size distribution, R_i is the median radius of the aerosols (μm), σ_i is the standard deviation, and C_i is the volume concentration of particulate matter ($\mu m^3/\mu m^2$). The subscripts f and c represent the fine and coarse modes, respectivaly. Additionally, the parameters in the fine and coarse modes are listed in Table 1.

Table 1. The parameters of the water-soluble (WS), biomass burning (BB), and dust (DU) models in the fine mode fraction (FMF) calculation.

Type	$r_f/\mu m$	$r_c/\mu m$	$ln\sigma_f$	$ln\sigma_c$	C_f/C_c	n	k	FMF_v
WS	0.118	1.17	0.6	0.6	2	1.45	0.0035	0.65
BB	0.132	4.50	0.4	0.6	4	1.52	0.025	0.80
DU	0.100	3.40	0.6	0.8	0.066	1.53	0.008	0.06

Note: n and k are the real part and imaginary part of complex refractive index, respectively.

Table 2 shows the FMF results of three different aerosol models based on two algorithms. We find that the FMFs in the same model are very close to each other. The maximum difference of FMF (0.04) is presented in the WS model, while the minimum difference in the DU model is only 0.01. It is suggested that the theoretical difference of FMF under the two algorithms can be ignored, and the FMF fusion can be executed between ground-based FMF and MODIS FMF.

Table 2. The FMF differences between the Moderate Resolution Imaging Spectroradiometer (MODIS) algorithm and the spectral deconvolution algorithm (SDA) method in the WS, BB, and DU models.

Type	FMF(MODIS)	FMF(SDA)	Error(MODIS-SDA)
WS	0.82	0.78	0.04
BB	0.98	0.95	0.03
DU	0.31	0.32	0.01

2.2. Fusion Method for Merging FMF

Our fusion method includes three steps: (i) analysis of the experimental variogram; (ii) estimated parameters in the experimental variogram; (iii) merging space-borne and ground-based FMF using the UK method. In our study, the UK method is employed to merge the FMF, which was proposed by G. Matheron in 1969 and has been widely used in several fields (e.g., mining, agriculture, environmental, and atmosphere sciences) [33–35]. The flowchart is shown in Figure 1.
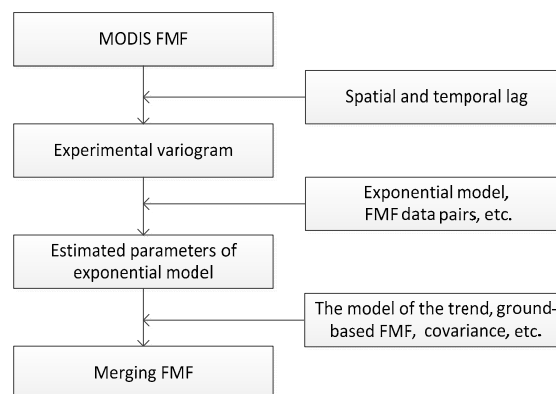
Figure 1. The flowchart of the fusion method applied to merging MODIS FMF and ground-based FMF.

(i)　Analysis of the experimental variogram:

For all of the FMF pairs at any two locations (times) from the MODIS products, the experimental variogram is evaluated as:

$$\gamma(h_x, h_t) = \frac{1}{2}\left[FMF(x_i, t_i) - FMF(x_j, t_j)\right]^2 \tag{4}$$

where $FMF(x_i, t_i)$ is the FMF observation at location x_i and time t_i; h_x is the spatial distance, which can be calculated by the longitudes and latitudes of the spatial points; and h_t is the time lag in days between two temporal points. Thus, h_x can be expressed as follows:

$$h_x(x_i, x_j) = Re \cdot \arccos\left(sin\varphi_i sin\varphi_j + cos\varphi_i cos\varphi_j \cos(\theta_i - \theta_j)\right) \tag{5}$$

where Re is the mean radius of the Earth, φ and θ are latitude and longitude separately, and x_i and x_j are the locations of the observations.

If the temporal changes are neglected, the exponential model can well represent the spatial autocorrelation and is often applied to fitting the experimental variogram:

$$\gamma_{theo}(h_x) = \begin{cases} 0, \ h_x = 0 \\ \sigma_n^2 + \sigma_b^2 exp(-h_x/l), \ h_x > 0 \end{cases} \tag{6}$$

where $\gamma_{theo}(h_x)$ represents the theoretical spatial semivariance at an h_x separation distance and $\sigma^2 (= \sigma_n^2 + \sigma_b^2)$ represents the variance of FMF observations beyond the correlation length (3 l). σ_n^2 is the nugget, representing both the measurement error and the micro-variability at distances smaller than one pixel. σ_b^2 represents the variance of the FMF in the correlation length (3 l). l is the range parameter.

(ii)　Estimated parameters in the experimental variogram:

A least squares method is employed to estimate the parameters in the experimental variogram (Equation (6)). Firstly, the FMF distribution from MODIS is applied to fit the initial variance and correlation length. Then, the MODIS FMFs are matched at the ground-based sites to generate the FMF pairs. Using the FMF pairs, the final parameters can be fit to the experimental variogram. Generally, a higher variance is indicated for greater overall variability, and a shorter correlation length is related to a greater spatial variability at smaller scales.

(iii)　Merging space-borne and ground-based FMF using the UK method:

When given the ground-based FMF measurements at n locations and times, the estimated FMF distributions **s** at m locations and times can be expressed as the sum of an unknown component $X_s\beta$ and a zero-mean stochastic component v:

$$\mathbf{s} = X_s\beta + v \tag{7}$$

where X_s ($m \times 2$) defines the model of the trend, and β is a 2 × 1 vector of drift coefficients that can define the weights of the two variables in the model of the trend. Since FMFs are measured by ground-based sensors as well as MODIS, a linear relationship between them is the simplest and most reasonable model, yielding a linear model of the trend:

$$X_s = \begin{bmatrix} 1 & FMF_{11} \\ \vdots & \vdots \\ 1 & FMF_{m1} \end{bmatrix} \tag{8}$$

The model of the trend includes a column of ones and a column of MODIS FMF at all estimated locations and times. The column of ones in the model of the trend will multiply β, representing

an overall constant. Analogous to a regression model, the constant term aims to obtain a net offset not captured by MODIS. At the ground-based (*n*) locations, the stochastic component ν represents the observed spatial and temporal residuals between the ground-based FMF measurements and the weighted MODIS FMF observations. At the estimation (*m*) locations, this component represents the predicted residuals between the true FMF and the weighted MODIS FMF at those locations/times. Further derivation, including the covariance of these residuals and UK equations, is described in Appendix A.

2.3. PMRS Model

PMRS, proposed by Zhang and Li (2015) [13], is a remote sensing model to estimate dry $PM_{2.5}$ mass concentration near the ground using the following formula:

$$PM_{2.5} = AOD \frac{FMF \cdot VE_f \cdot \rho_{f,dry}}{PBLH \cdot f_o(RH)} \tag{9}$$

where $\rho_{f,dry}$ is the density of dry fine particulate matter, *PBLH* is the planetary boundary layer height, *RH* is the relative humidity, and VE_f is the columnar volume-to-extinction ratio, which can be obtained using *FMF*; f_o is the particle drying factor depending on *RH*. VE_f can be expressed as follows:

$$VE_f = 0.2887\, FMF^2 - 0.4663 FMF + 0.356 \quad (0.1 \le FMF \le 1.0) \tag{10}$$

2.4. Data and Study Area

The AOD and FMF parameters can be retrieved over land and ocean using the dark target method [36] from MODIS (https://MODIS.gsfc.nasa.gov/) onboard Terra and Aqua. In this study, the Terra Collection 006 FMF (Optical_Depth_Ratio_Small_Land) is used to merge with ground-based data and merging FMF and AOD (Optical_depth_Land_and_Ocean) to calculate $PM_{2.5}$ with the PMRS model. The spatial resolution of AOD and FMF is 10 km.

For ground-based observations, both AERONET data (Version 2) and SONET data are used. AERONET is a global ground-based aerosol monitoring network that offers easy access to the data archive, and, currently, it has more than 600 sites [37,38] (http://aeronet.gsfc.nasa.gov/) but only six permanent (more than five years) sites in study area of Eastern China (108° E to 123° E, 28° N to 43° N) (Figure 2). SONET has been operational by the Chinese Academy of Sciences since 2010. There are an additional six permanent sites distributed in the study area [39–41]. Combining ground-based sites in these two networks (AERONET and SONET), the volume of FMF data can be effectively increased in the study area. Furthermore, the FMF from SONET is retrieved by the SDA method in the same way as AERONET. We choose to use both Lev 2.0 (i.e., Beijing-CAMS and Xianghe) and 1.5 data (other sites) of AERONET due to difficulties in calibration and maintenance in China. For SONET, we use the Lev 1.6 data, which reflect the current optimum data-set published at http://www.sonet.ac.cn.

In addition, the auxiliary data of RH and PBLH are extracted from National Centers for Environmental Prediction Final Operational Global Analysis data (NCEP FNL) (http://rda.ucar.edu/), which has a temporal resolution of 6 h and a spatial resolution of 1° × 1° [42]. The in situ $PM_{2.5}$ measurements from the Ministry of Environmental Protection (MEP), China, are used to validate the $PM_{2.5}$ estimates from the PMRS model using FMF fusion products.

The satellite and auxiliary data are resampled into 0.2° × 0.2° grid cells, applying bilinear interpolation to RH and PBLH data and area averaging to the MODIS observations. The FMF data from MODIS and ground-based sites cannot be compared directly since they have different spatial and temporal resolutions. Therefore, the mean of MODIS FMF observations within a 0.2° × 0.2° bounding box around each ground-based site is used to compare them with the ground-based FMF data, which are themselves averaged within ±30 min from the Terra overpass. According to the spatiotemporal analysis (See in Section 3.1), the FMF fusion is performed in one week increments, using weekly averaged ground-based and MODIS FMF data. For each seven day period, the ground-based sites that

have FMF data for at least three of the seven days and that have overlapping MODIS FMF data are used in this study. Two to five sites are used during various weeks.
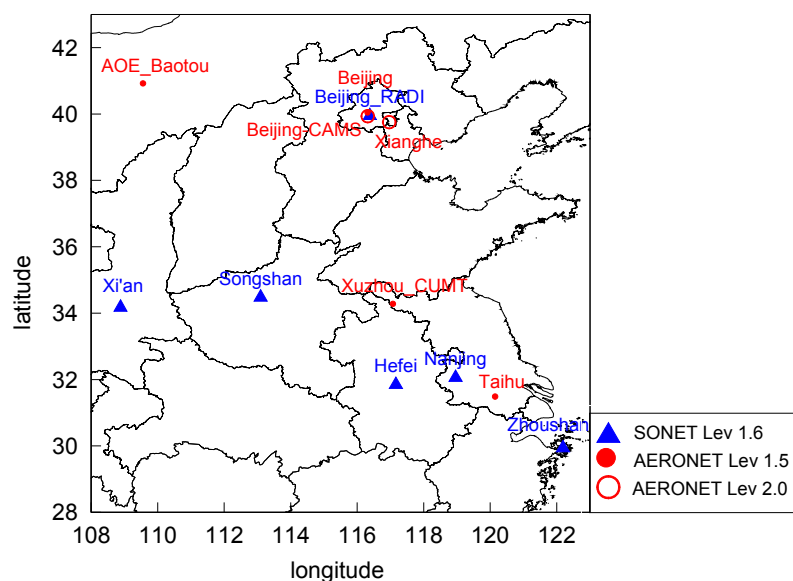
**Figure 2.** The study area and ground-based observation locations used in the study.

3. Results

3.1. Spatiotemporal Variability Analysis

To analyze the spatio-temporal variability of FMF, we calculate the raw variogram of each 16-day cycle from December 2014 to December 2015 using all of the data in the study area. Then, the raw variogram is separated by a spatial lag of 100 km and a temporal lag of one day. Figure 3a shows the spatial and temporal variograms for all of the MODIS data from 19 January to 3 February 2015. The horizontal axis represents the spatial distance between two observations, while the vertical axis represents the time difference in which the two observations are recorded. In the time scale, the figure does not show any noticeable temporal autocorrelation in the day-to-day variability of FMF distribution at the same distance lag up to seven days. This is likely a result of the change in FMF distribution over these days or of the fact that MODIS does not acquire a change in the FMF for a short time in the area. Thus, the FMF fusion is performed in one week increments. Furthermore, the spatial autocorrelation is obvious since the variances increase with the spatial distance. To show the spatial autocorrelation feature, all of the MODIS FMF data from 19 January to 3 February 2015 are examined in a single spatial variogram (Figure 3a,b). The maximum range of distance is set to 1000 km since the number of pairs of observations in each bin becomes fewer as the spatial distance becomes longer. Figure 3b shows the spatial autocorrelation feature. We can find that the exponential model is well fit to the experimental variogram. In this example case, the spatial correlation of FMF data is approximately 1425 km, and the spatial distance beyond this value is independent.

Figure 3c presents the parameters of the fitted theoretical spatial variograms for each repeat cycle from December 2014 to December 2015. The smaller correlation lengths during winter and spring indicate the significantly heterogeneous distribution of FMF, mainly because of the frequent occurrence of regional haze in Winter and frequent sandstorms in Spring in the study area. In Summer, the correlation lengths are abnormal (close to zero or larger than 10 hundred kilometers), which fail to fit reasonable parameters. The variance of the FMF observations over a long distance (beyond the correlation distance) ranges from 0.05 to 0.19 throughout the study time period. The variance is slightly higher in Winter and Spring corresponding to the relatively smaller correlation length. As a result, the

fusion can be performed in one-week increments, and the spatial-only experimental variogram is in good agreement with the exponential model based on MODIS products.
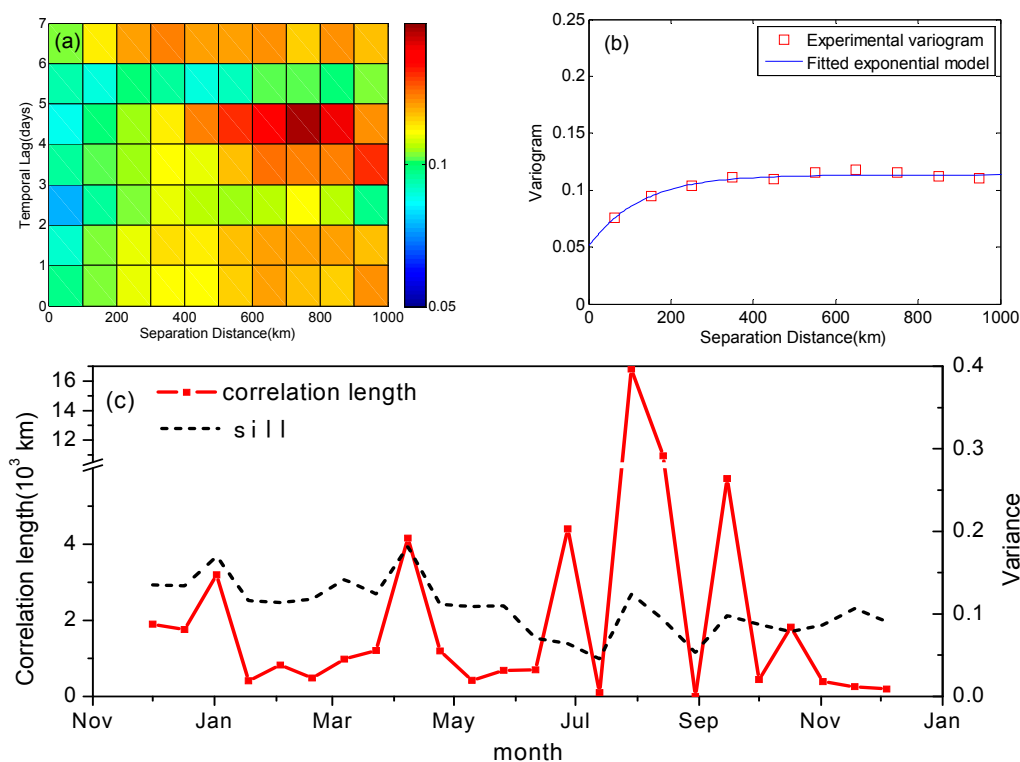
Figure 3. (**a**) The spatial and temporal variograms of FMF over a 16-day period from 18 January to 2 February 2015 for MODIS. The color bar indicates the semivariance; (**b**) The spatial variogram over the same period from MODIS; (**c**) Correlation length, variance of FMF for all 16-day periods from December 2014 to December 2015.

We use the parameters in the exponential model from MODIS products as the initial value of the variogram analysis to match a group of parameters in the experimental variogram using both MODIS and ground-based FMF in the Winter of 2015. The nugget value (σ_n^2.) is found to be zero (0.0018); σ_b^2 is 0.0141; and the range parameter l is 475 km.
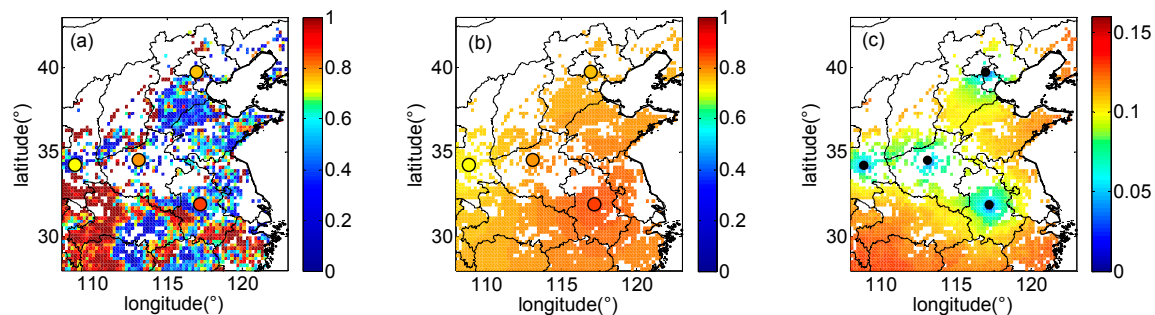
3.2. Fusion Results

3.2.1. FMF Test Case

Figure 4a shows the mean MODIS FMF observations from 22 December to 28 December 2014. Note that the white gaps mean no observations in this period. It shows that the MODIS FMF observations vary from 0.1 to one and the discrepancy of FMF observations next to each other may reach up to 0.6. Higher FMF mainly is distributed in the Southwest of the study area. At ground-based locations, the mean absolute error between MODIS FMF and ground-based observations is 0.48, and the maximum error reaches up to 0.64. The FMF fusion result is shown in Figure 4b for the same period. The FMF fusion results mainly range from 0.60 to 0.90, and higher values are distributed in the southern regions in a similar fashion to the ground-based FMF observations. The mean absolute error at ground-based sites decreases to 0.002. The predicted uncertainties shown in Figure 4c are to evaluate the efficiency of this method. The lowest uncertainty with a value of about 0.03 is presented at the ground-based locations, while it is higher within the correlation length and the highest beyond the correlation length with no more than 0.15. The uncertainties are deemed to be within acceptable limits. In addition, the propagating error of estimated FMFs derived from ground-based FMF is relatively

larger than that from MODIS. However, the ground-based FMF has a high degree of accuracy, so the errors are minimized.

Figure 4. (**a**) Mean FMF observations from MODIS; (**b**) FMF fusion result obtained from universal kriging (UK); and (**c**) uncertainty associated from the UK estimates, expressed as a standard deviation. All data are for the period from 22 December to 28 December 2014. The colored solid circles in the figures are the ground-based locations and their values.

3.2.2. Comparison and Validation

Before merging the FMF data, a correlation analysis is performed between ground-based FMF and MODIS FMF in the Winter of 2015 to evaluate their temporal and spatial consistency. According to the principle of the spatial and temporal matching of data (see Section 2.4), we can obtain the matching data pairs, as shown in Figure 5. MODIS FMF data have a wide range from 0.1 to one (the value of zero is discarded), while ground-based FMF values are mainly concentrated from 0.6 to 0.9. The maximum absolute error (AE) between them is about 0.75. The correlation coefficient is very low mainly because of the low accuracy of MODIS FMF data. It should be noted that approximately 30 percent of the data pairs have a high consistency (AE < 0.2), indicating that the algorithm of MODIS FMF is not stable when retrieving FMF. In some cases, reasonable FMF retrievals can be obtained. The situation of significant error between data pairs can be divided into two types: (1) the highest value of 1.0, which is overestimated, and (2) significant underestimation in other cases. This may be due to improper selection of the aerosol types in the MODIS FMF retrieval algorithm.
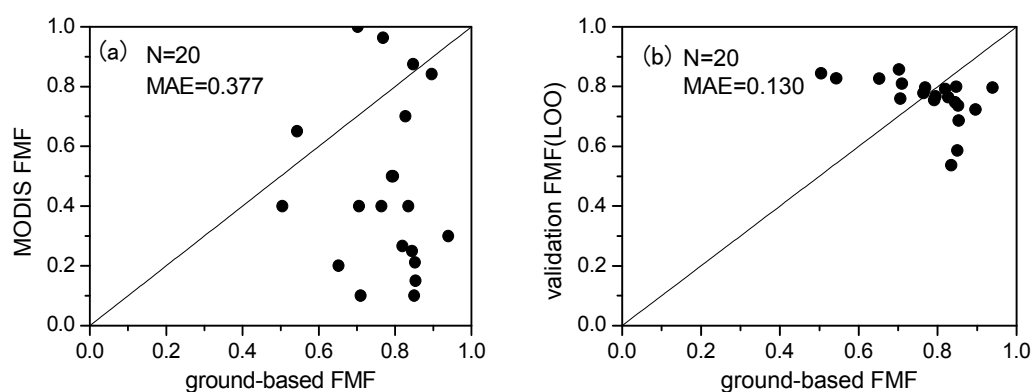
Figure 5. (**a**) Matching data pairs for MODIS FMF and ground-based FMF (*N* = 20); and (**b**) leave-one-out cross-validation results (*N* = 20).

To validate and evaluate the efficiency of the fusion approach of UK, leave-one-out (LOO) cross-validation is used to validate the FMF fusion results in the Winter of 2015. The validation method is to select any one of the ground-based locations as unknown and to use the other ground-based locations as known to merge with MODIS FMF data. The validation results can be obtained by comparing the fusion results at ground-based locations with their measurements.

Comparing the pre-fusion data pairs and leave-one-out validation results (Table 3), we can find that MODIS FMF data have a wide range from 0.1 to one (the value of zero is discarded), while ground-based FMF values are mainly concentrated from 0.50 to 0.95. The maximum absolute error (AE) between them is 0.75, which is a great error for FMF with an interval of zero to one. After fusion, the maximum absolute error is decreased to 0.34, which is significantly lower than that before fusion (0.75). The point with a big error (0.34) is because of a relative high offset in the fusion. In addition, the mean absolute error between MODIS FMF and ground-based FMF is 0.38, which is much higher than the 0.13 after fusion. In conclusion, the accuracy of FMF has been remarkably improved and is much more reliable after fusion.

Table 3. Comparison between the pre-fusion and leave-one-out validation results of Winter 2015 at ground-based locations [a].

Value	Ground-Based FMF	MODIS FMF	FMF (LOO)	Δ (MODIS)	Δ (LOO)
Maximum	0.94	1.00	0.86	0.75	0.34
Minimum	0.50	0.10	0.54	0.03	0.01
Mean	0.77	0.46	0.76	0.38	0.13

[a] Ground-based FMF is FMF from the Aerosol Robotic Network (AERONET) and the Sun-sky radiometer Observation Network (SONET); MODIS FMF is FMF from MODIS; FMF (leave-one-out (LOO)) is the cross-validation results of FMF; Δ (MODIS) and Δ (LOO) are the errors between ground-based FMF and MODIS FMF and between ground-based FMF and FMF (LOO), respectively.

3.3. $PM_{2.5}$ Results

### 3.3.1. $PM_{2.5}$ Test Case

To give a description of the effectiveness of the FMF fusion results in deriving $PM_{2.5}$ near the surface, the MODIS FMF and the FMF fusion results are used as the input parameters of the PMRS model separately. Figure 6 illustrates not only the two kinds of $PM_{2.5}$ mass concentration results, but also the $PM_{2.5}$ measurements from the MEP for a period from 9 February to 15 February 2015. In Figure 6a, $PM_{2.5}$ measurements in the whole study area are below 200 μg/m^3, and most of them are no more than 100 μg/m^3. Higher $PM_{2.5}$ mass concentration is mainly distributed in the Southwest of the study area, with relatively lower values in other regions. As for Figure 6b, higher values are distributed not only in the North, but also in the Southwest of the study area, which is more consistent with the in situ $PM_{2.5}$ measurements. In the North of the study area, a point located in Shandong Province has a large discrepancy compared with $PM_{2.5}$ from the MEP mainly because of a lack of sampling or a failure to observe the pollution process by the satellite. In conclusion, $PM_{2.5}$ from FMF fusion results has better consistency with the measurements from the MEP.
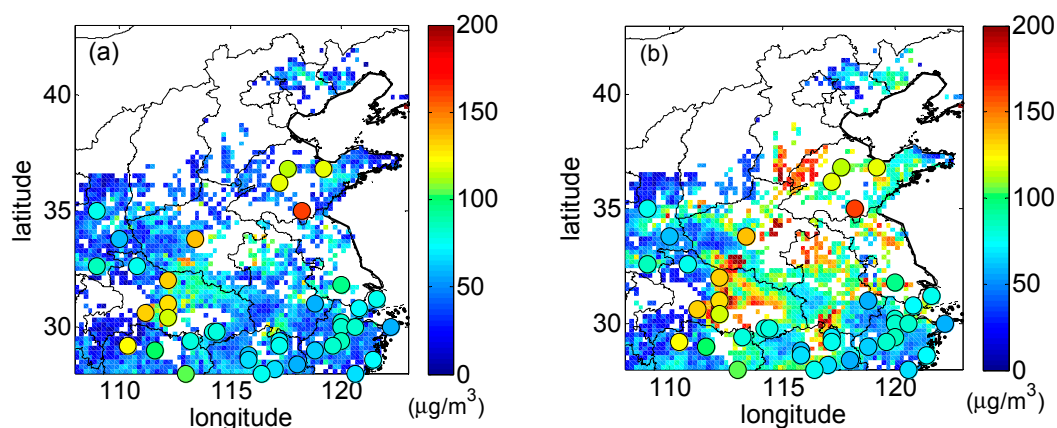
Figure 6. (**a**) PM$_{2.5}$ (atmospheric particulate matter with a mass median diameter less than 2.5 μm) mass concentration near the ground estimated from FMF observations; and (**b**) FMF fusion results based on UK method for a time period from 9 February to 15 February 2015. The colored solid circles in the figures are PM$_{2.5}$ measurements from the Ministry of Environmental Protection (MEP). The white gaps indicate no values.

3.3.2. Validation of PM$_{2.5}$

To quantify the influence of merging FMF on estimating PM$_{2.5}$, validations are performed by making comparisons between the PM$_{2.5}$ estimated from fusion FMF, MODIS FMF observations, and PM$_{2.5}$ in situ measurements in the Winter of 2015. The results are shown in Figure 7. PM$_{2.5}$ estimated from MODIS observations and FMF fusion results is lower than the PM$_{2.5}$ in situ measurements. Since the PM$_{2.5}$ monitoring stations are mainly distributed in more polluted cities rather than rural areas, the in situ measurements will be higher than remote sensing estimates. Furthermore, the PM$_{2.5}$ mass concentration is generally high during haze days, but the satellite is limited to retrieve AOD and FMF successfully on highly polluted days. Thus, it fails to obtain some high PM$_{2.5}$ mass concentrations. Compared with the mean of in situ PM$_{2.5}$ measurements (88.9 μg/m$^{3)}$, the mean of PM$_{2.5}$ estimated from FMF fusion results is much closer than that of PM$_{2.5}$ estimated from MODIS FMF (87.2 μg/m^3 vs. 49.8 μg/m^3). Additionally, the mean error of PM$_{2.5}$ estimated from FMF fusion results is smaller than that of PM$_{2.5}$ estimated from MODIS FMF (41.8 μg/m^3 vs. 45.3 μg/m^3). In conclusion, the FMF fusion result can provide FMF of higher accuracy as the input parameter of PMRS, thus having a better understanding of PM$_{2.5}$.
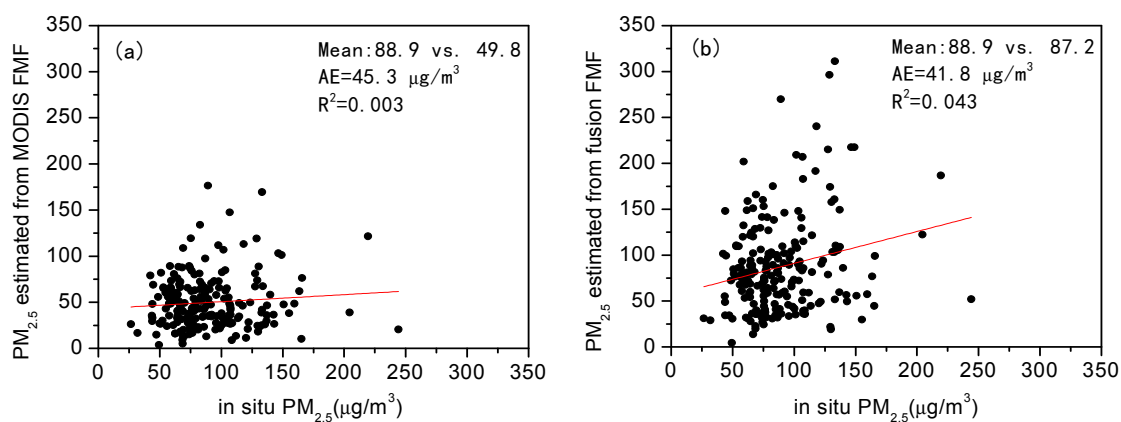

Figure 7. (**a**) Comparison between in situ PM$_{2.5}$ and PM$_{2.5}$ estimated from MODIS FMF; (**b**) Comparison between in situ PM$_{2.5}$ and PM$_{2.5}$ estimated from FMF fusion results. Note that the red lines are the fitted regression lines.

It should be noted that the PMRS model is a semi-physical model and has several key parameters. Improving FMF based on the UK method can effectively correct the system error of the model. However, the error that caused by the coarse spatial resolution of RH and PBLH cannot be corrected currently.

3.4. Spatio-Temporal Characteristics Analysis of FMF and PM$_{2.5}$

To have a better understanding of the spatio-temporal feature of FMF and PM$_{2.5}$ in winter, the mean FMF and PM$_{2.5}$ results from fusion FMF are shown in Figure 8. The mean FMF in the study area is mainly within the 0.6 to 0.9 distribution, and FMF in the North (mainly in Beijing and its surrounding area) and the middle of study area is relatively high compared with the surrounding area. As for PM$_{2.5}$, most of the PM$_{2.5}$ mass concentration is less than 200 μg/ m^3 in the study area, and higher PM$_{2.5}$ is in the middle and South of the study area. Thus, FMF and PM$_{2.5}$ have no inevitable connection. To obtain variability in Winter, we choose three regions (8 × 8 pixels) and calculate the mean FMF and PM$_{2.5}$ to have a detailed analysis. The regions are marked in Figure 8a. The FMF in Region A is relatively high. As for Regions B and C, a higher PM$_{2.5}$ mass concentration is shown in Figure 8b.
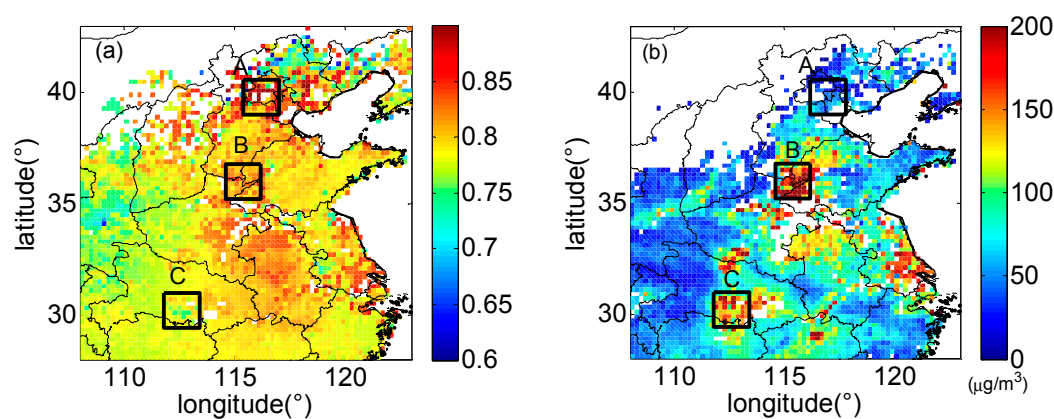
Figure 8. (**a**) The mean FMF fusion results in the Winter of 2015 and the three regions in black squares; and (**b**) the mean PM$_{2.5}$ estimated from the FMF fusion results in the Winter of 2015 (there is no result in the time period from 26 January 2015 to 1 February 2015 because of the lack of data pairs).

The variabilities of FMF and PM$_{2.5}$ in each region in Winter are shown in Figure 9. We can see that FMF in each region changes slightly, while PM$_{2.5}$ is rather changeable. Region A, which mainly consists of Beijing, Tianjin, and part of Hebei, has a low consistency of the variation trends not only between PM$_{2.5}$ in situ measurements and PM$_{2.5}$ estimated from fusion FMF, but also between the ground-based FMF measurements and FMF fusion results. This is mainly because of a lack of samplings resulting from the high surface albedos of this area in winter. Region C has the highest consistency of PM$_{2.5}$ from in situ measurements and estimated from FMF fusion results. In Regions A and B, the averages between in situ PM$_{2.5}$ measurements and PM$_{2.5}$ estimates in the former six time periods are close to each other. In the last several time periods (namely 12 January 2015 to 28 February 2015), the errors between measurements and estimates in Regions A and B are higher, with a maximum value of 150 μg/m^3, probably because the PM$_{2.5}$ sites mainly are distributed in cities. As a whole, PM$_{2.5}$ estimated from FMF fusion results is lower than from MEP since the sites of MEP mainly are distributed in cities.
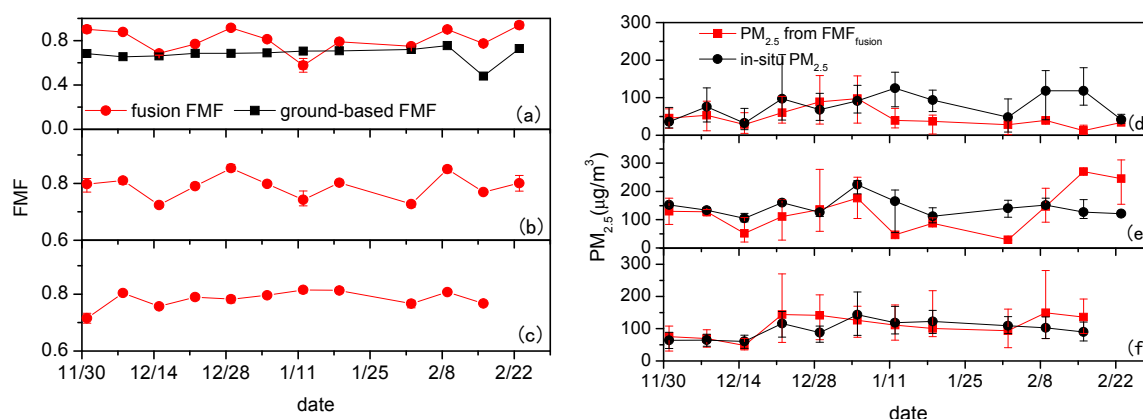
Figure 9. The variabilities of fusion results and ground-based FMF and variabilities of $PM_{2.5}$ from in-situ measurements and estimated from fusion results in each region, including the mean, maximum, and minimum value in each region. (**a**,**d**) are from Region A; (**b**,**e**) are from Region B; (**c**,**f**) are from Region C.

4. Discussion and Conclusions

In this paper, a UK method is applied to the fusion of MODIS FMF data and ground-based FMF data over eastern China in the Winter of 2015. To test the applicability of this method, the fusion results have been validated using leave-one-out cross-validation and applied to derive $PM_{2.5}$ using a PMRS model. The validation results show that the mean error of fusion FMF decreases from 0.38 to 0.13, with the maximum error of 0.34 corresponding to that of 0.75 before data fusion. $PM_{2.5}$ estimated from FMF fusion results is closer to $PM_{2.5}$ from MEP. The mean error of $PM_{2.5}$ estimated from fusion FMF is reduced by about 1.7 $\mu g/m^3$ (87.2 $\mu g/m^3$ vs 88.9 $\mu g/m^3$) compared with $PM_{2.5}$ from MEP. In conclusion, merging FMF using the UK method provides the input FMF parameter of the PMRS model with high accuracy, thus providing support for air quality monitoring.

Since there are limited AERONET sites that have Lev 2.0 data and the method requires as many ground reference data as possible, we have used Lev 1.5 data as a supplement. The wrong FMF value, which is caused by cloud, cannot be used due to the combination with MODIS.

FMF data can be carried out on instantaneous fusion in the future to provide reliable data for short-term study such as weather research and air quality monitoring. With the FMF data of multi-sensors being available and reliable, FMF fusion can give better estimates and further our understanding of FMF and fine particulate matters.

Author Contributions: Aimei Zhao designed the experiment, performed the fusion method, and wrote this paper. Ying Zhang and Zhengqiang Li conceived of the experiment and revised the paper. Yang Zhang provided technical guidance and revised the paper. Donghui Li assisted with the ground-based data collection and provided useful advice regarding the data.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

A.1. UK Theory in Merging FMF

In this case, the variogram analysis is carried out on the detrended ground-based FMF data ($\mathbf{g} - \mathbf{X_g}\boldsymbol{\beta}$, $\mathbf{g}$ is the ground-based FMF data, $\mathbf{X_g}$ is the trend model at ground-based sites, and $\boldsymbol{\beta}$ is the drift coefficient) within the study area. The detrending is carried out by subtracting the trend composed by the constant plus the MODIS FMF ($\mathbf{X_g}$) by their corresponding drift coefficients. The nugget σ_n^2 in this case is found to be zero. Since the nugget represents both measurement error and microvariability, this indicates that the measurement error of ground-based FMF is negligible relative to the total amount of variability.

The covariance of these residuals is described using a matrix $\mathbf{Q}$, wherein the covariance function is defined based on the variogram analysis (Equation (A1)) such that the covariance between two points x_i and x_j is defined as:

$$Q_{ij} = \sigma^2 - \gamma_{\text{theo}}(h_x) = \begin{cases} \sigma_n^2 + \sigma_b^2, \ h_x = 0 \\ \sigma_b^2 \exp(-h_x/l), \ h_x > 0 \end{cases} \tag{A1}$$

The estimates of FMF at target times and locations are obtained by solving the following system of universal kriging equations:

$$\begin{bmatrix} \mathbf{Q_{gg}} & \mathbf{X_g} \\ \mathbf{X_g^T} & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda}^{\mathrm{T}} \\ \mathbf{M} \end{bmatrix} = \begin{bmatrix} \mathbf{Q_{gs}} \\ \mathbf{X_s^T} \end{bmatrix} \tag{A2}$$

where $\mathbf{Q_{gs}}$ ($n \times m$) is defined as the spatial covariance matrix between ground-based locations' observations and the estimate locations based on Equation (A1). $\mathbf{X_s}$ ($m \times 2$) is the model of the trend defined at the estimate locations based on Equation (8). Solve the system of equations to obtain $\boldsymbol{\Lambda}$ ($m \times n$), which is the weight to be assigned to ground-based FMF observations for interpolation to the m locations and times, and M is a $2 \times m$ matrix of Lagrange multipliers. The FMF distribution $\widetilde{\mathbf{s}}$ and the uncertainty $\mathbf{V_{\widetilde{s}}}$ associated with the estimates are as follows:

$$\widetilde{\mathbf{s}} = \boldsymbol{\Lambda} g \tag{A3}$$

where $\widetilde{\mathbf{s}}$ is the final FMF estimate at m locations/ times that includes information derived from all of the sensors, and $\mathbf{g}$ is the n ground-based observations. The uncertainty associated with these estimates is:

$$\mathbf{V_{\widetilde{s}}} = \mathbf{Q_{gs}} - \boldsymbol{\Lambda}^{\mathrm{T}}\mathbf{Q_{gs}} - \mathbf{X_s M} \tag{A4}$$

where the diagonal elements of $\mathbf{V_{\widetilde{s}}}$ represent the predicted error variance ($\sigma_{\widetilde{s}}^2$) of individual elements in $\widetilde{\mathbf{s}}$, i.e., of the estimates of FMF at individual locations and times. Finally, the influence of MODIS on the final estimates $\widetilde{\mathbf{s}}$ can be quantified by calculating the estimated drift coefficients $\boldsymbol{\beta}$ (2×1) and the corresponding uncertainties for the data set:

$$\boldsymbol{\beta} = \left(\mathbf{X_g^T Q_{gg}^{-1} X_g}\right)^{-1} \mathbf{X_g^T Q_{gg}^{-1} g} \tag{A5}$$

$$\mathbf{V_{\widetilde{\beta}}} = \left(\mathbf{X_g^T Q_{gg}^{-1} X_g}\right)^{-1} \tag{A6}$$

where the diagonal elements of $\mathbf{V_{\widetilde{\beta}}}$ represent the uncertainty of the individual drift parameters ($\sigma_{\widetilde{\beta}}^2$) and the off-diagonal terms represent the estimated covariance of the errors associated with these estimates.

Using these drift coefficients, the estimated FMF field can also be expressed as:

$$\hat{\mathbf{s}} = \mathbf{X_s}\hat{\boldsymbol{\beta}} + \mathbf{Q_{gs}^T Q_{gg}^{-1}}(\mathbf{g} - \mathbf{X_z}\hat{\boldsymbol{\beta}}) \tag{A7}$$

References

1. Pope, C.A.; Burnett, R.T.; Thun, M.J.; Calle, E.E.; Krewski, D.; Ito, K.; Thurston, G.D. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *JAMA* **2002**, *287*, 1132–1141. [CrossRef]

2. Kunzli, N.; Jerrett, M.; Mack, W.J.; Beckerman, B.; LaBree, L.; Gilliland, F.; Thomas, D.; Peters, J.; Hodis, H.N. Ambient air pollution and atherosclerosis in Los Angeles. *Environ. Health Perspect.* **2005**, *113*, 201–206. [CrossRef] [PubMed]

3. Brook, R.D.; Rajagopalan, S.; Pope, C.A.; Brook, J.R.; Bhatnagar, A.; Diez-Roux, A.V.; Holguin, F.; Hong, Y.L.; Luepker, R.V.; Mittleman, M.A.; et al. Particulate matter air pollution and cardiovascular disease an update to the scientific statement from the American heart association. *Circulation* **2010**, *121*, 2331–2378. [CrossRef] [PubMed]

4. Liu, Y.; Paciorek, C.J.; Koutrakis, P. Estimating regional spatial and temporal variability of PM$_{2.5}$ concentrations using satellite data, meteorology, and land use information. *Environ. Health Perspect.* **2009**, *117*, 886–892. [CrossRef] [PubMed]

5. Paciorek, C.J.; Liu, Y. Limitations of remotely sensed aerosol as a spatial proxy for fine particulate matter. *Environ. Health Perspect.* **2009**, *117*, 904–909. [CrossRef] [PubMed]

6. Wang, J.; Christopher, S.A. Intercomparison between satellite-derived aerosol optical thickness and PM$_{2.5}$ mass: Implications for air quality studies. *Geophys. Res. Lett.* **2003**, *30*, 4. [CrossRef]

7. Lee, H.J.; Coull, B.A.; Bell, M.L.; Koutrakis, P. Use of satellite-based aerosol optical depth and spatial clustering to predict ambient PM$_{2.5}$ concentrations. *Environ. Res.* **2012**, *118*, 8–15. [CrossRef] [PubMed]

8. Hutchison, K.D.; Smith, S.; Faruqui, S.J. Correlating MODIS aerosol optical thickness data with ground-based PM$_{2.5}$ observations across Texas for use in a real-time air quality prediction system. *Atmos. Environ.* **2005**, *39*, 7190–7203. [CrossRef]

9. Engel-Cox, J.A.; Young, G.S.; Hoff, R.M. Application of satellite remote-sensing data for source analysis of fine particulate matter transport events. *J. Air Waste Manag. Assoc.* **2005**, *55*, 1389–1397. [CrossRef] [PubMed]

10. Drury, E.; Jacob, D.J.; Spurr, R.J.D.; Wang, J.; Shinozuka, Y.; Anderson, B.E.; Clarke, A.D.; Dibb, J.; McNaughton, C.; Weber, R. Synthesis of satellite (MODIS), aircraft (ICARTT), and surface (IMPROVE, EPA-AQS, AERONET) aerosol observations over eastern north America to improve MODIS aerosol retrievals and constrain surface aerosol concentrations and sources. *J. Geophys. Res. Atmos.* **2010**, *115*, 17. [CrossRef]

11. Wang, J.; Xu, X.G.; Spurr, R.; Wang, Y.X.; Drury, E. Improved algorithm for MODIS satellite retrievals of aerosol optical thickness over land in dusty atmosphere: Implications for air quality monitoring in China. *Remote Sens. Environ.* **2010**, *114*, 2575–2583. [CrossRef]

12. Xu, X.G.; Wang, J.; Henze, D.K.; Qu, W.J.; Kopacz, M. Constraints on aerosol sources using GEOS-Chem adjoint and MODIS radiances, and evaluation with multisensor (OMI, MISR) data. *J. Geophys. Res. Atmos.* **2013**, *118*, 6396–6413. [CrossRef]

13. Zhang, Y.; Li, Z.Q. Remote sensing of atmospheric fine particulate matter (PM$_{2.5}$) mass concentration near the ground from satellite observation. *Remote Sens. Environ.* **2015**, *160*, 252–262. [CrossRef]

14. Xu, H.; Xue, Y.; Guang, J.; Li, Y.; Yang, L.; Hou, T.; He, X.; Dong, J.; Chen, Z. A semi-empirical optical data fusion technique for merging aerosol optical depth over China. In Proceedings of the 2012 IEEE International Geoscience and Remote Sensing Symposium, Munich, Germany, 22–27 July 2012; pp. 2524–2527.

15. Loeb, N.G.; Sun, W.; Miller, W.F.; Loukachine, K.; Davies, R. Fusion of CERES, MISR, and MODIS measurements for top-of-atmosphere radiative flux validation. *J. Geophys. Res. Atmos.* **2006**, *111*, D18209. [CrossRef]

16. Singh, M.K.; Venkatachalam, P. Merging of aerosol optical depth data from multiple remote sensing sensors. In Proceedings of the 2014 IEEE Geoscience and Remote Sensing Symposium, Quebec City, QC, Canada, 13–18 July 2014; pp. 4173–4175.

17. Xu, Q.; Obradovic, Z.; Han, B.; Li, Y.; Braverman, A.; Vucetic, S. Improving Aerosol Retrieval Accuracy by Integrating AERONET, MISR and MODIS Data. In Proceedings of the 2015 7th International Conference on Information Fusion, Philadelphia, PA, USA, 25–28 July 2005.

18. Chatterjee, A.; Michalak, A.M.; Kahn, R.A.; Paradise, S.R.; Braverman, A.J.; Miller, C.E. A geostatistical data fusion technique for merging remote sensing and ground-based observations of aerosol optical thickness. *J. Geophys. Res. Atmos.* **2010**, *115*, 12. [CrossRef]

19. Jinnagara Puttaswamy, S.; Nguyen, H.M.; Braverman, A.; Hu, X.; Liu, Y. Statistical data fusion of multi-sensor AOD over the continental United States. *Geocarto Int.* **2014**, *29*, 48–64. [CrossRef]

20. Guo, J.; Gu, X.; Yu, T.; Cheng, T.; Chen, H.; Xie, D. *Trend Analysis of the Aerosol Optical Depth over China Using Fusion of MODIS and MISR Aerosol Products via Adaptive Weighted Estimate Algorithm*; IOP Publishing Ltd.: Bristol, UK, 2013; p. 88661X.

21. Xu, H.; Guang, J.; Xue, Y.; de Leeuw, G.; Che, Y.H.; Guo, J.; He, X.W.; Wang, T.K. A consistent aerosol optical depth (AOD) dataset over mainland China by integration of several aod products. *Atmos. Environ.* **2015**, *114*, 48–56. [CrossRef]

22. Leptoukh, G.; Zubko, V.; Gopalan, A. Spatial Aspects of Multi-Sensor Data Fusion: Aerosol Optical Thickness. In Proceedings of the 2007 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Barcelona, Spain, 23–28 July 2007; Sensing and Understanding Our Planet; IEEE: New York, NY, USA, 2007; Volumes 1–12, pp. 3119–3122.

23. Nirala, M. Technical note: Multi-sensor data fusion of aerosol optical thickness. *Int. J. Remote Sens.* **2008**, *29*, 2127–2136. [CrossRef]

24. Zubko, V.; Leptoukh, G.G.; Gopalan, A. Study of data-merging and interpolation methods for use in an interactive online analysis system: MODIS Terra and Aqua daily aerosol case. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 4219–4235. [CrossRef]

25. Li, L.; Shi, R.; Zhang, L.; Zhang, J.; Gao, W. The Data Fusion of Aerosol Optical Thickness Using Universal Kriging and Stepwise Regression in East China. In *Proceedings of SPIE;* International Society for Optics and Photonics: San Diego, CA, USA, 2014; p. 922112.

26. Nguyen, H.; Cressie, N.; Braverman, A. Spatial statistical data fusion for remote sensing applications. *J. Am. Stat. Assoc.* **2012**, *107*, 1004–1018. [CrossRef]

27. Yang, G.; Zhang, G.; Li, S. Application of universal kriging interpolation in geomagnetic map. *J. Chin. Inert. Technol.* **2008**, *8*, 162–166.

28. O'Neill, N.T.; Dubovik, O.; Eck, T.F. Modified Ångström exponent for the characterization of submicrometer aerosols. *Appl. Opt.* **2001**, *40*, 2368–2375. [CrossRef] [PubMed]

29. O'Neill, N.T.; Eck, T.F.; Holben, B.N.; Smirnov, A.; Dubovik, O.; Royer, A. Bimodal size distribution influences on the variation of Ångström derivatives in spectral and optical depth space. *J. Geophys. Res. Atmos.* **2001**, *106*, 9787–9806. [CrossRef]

30. O'Neill, N.T.; Eck, T.F.; Holben, B.N.; Smirnov, A.; Royer, A.; Li, Z. Optical properties of boreal forest fire smoke derived from sun photometry. *J. Geophys. Res. Atmos.* **2002**, *107*, 21. [CrossRef]

31. O'Neill, N.T.; Eck, T.F.; Smirnov, A.; Holben, B.N.; Thulasiraman, S. Spectral discrimination of coarse and fine mode optical depth. *J. Geophys. Res. Atmos.* **2003**, *108*, 15. [CrossRef]

32. Zhang, Y.; Li, Z.Q.; Wang, Y.; Li, K.T.; Li, D.H.; Zhang, Y.H.; Wei, P.; Wang, L.; Lu, Y. Improving accumulation-mode fraction based on spectral aerosol optical depth in Beijing. *Spectrosc. Spectr. Anal.* **2013**, *33*, 2795–2802.

33. Armstrong, M. Problems with universal kriging. *J. Int. Assoc. Math. Geol.* **1984**, *16*, 101–108. [CrossRef]

34. Stein, A.; Corsten, L.C.A. Universal kriging and cokriging as a regression procedure. *Biometrics* **1991**, *47*, 575–587. [CrossRef]

35. Yan, H. The Development and Implementation of Spatial Interpolation Technology. Master's Thesis, Southwest Petroleum University, Chengdu, China, April 2005.

36. Levy, R.C.; Remer, L.A.; Tanré, D.; Mattoo, S.; Kaufman, Y.J. Algorithm for Remote Sensing of Tropospheric Aerosol over Dark Targets from MODIS: Collections 005 and 051: Revision 2. 2009. Available online: http://modis.gsfc.nasa.gov/data/atbd/atbd_mod02.pdf (accessed on 3 July 2017).

37. Holben, B.N.; Eck, T.F.; Slutsker, I.; Tanre, D.; Buis, J.P.; Setzer, A.; Vermote, E.; Reagan, J.A.; Kaufman, Y.J.; Nakajima, T.; et al. AERONET—A federated instrument network and data archive for aerosol characterization. *Remote Sens. Environ.* **1998**, *66*, 1–16. [CrossRef]

38. Holben, B.N.; Tanre, D.; Smirnov, A.; Eck, T.F.; Slutsker, I.; Abuhassan, N.; Newcomb, W.W.; Schafer, J.S.; Chatenet, B.; Lavenu, F.; et al. An emerging ground-based aerosol climatology: Aerosol optical depth from AERONET. *J. Geophys. Res. Atmos.* **2001**, *106*, 12067–12097. [CrossRef]

39. Li, Z.; Li, D.; Li, K.; Xu, H.; Chen, X.; Chen, C.; Xie, Y.; Li, L.; Li, L.; Li, W. A sun/sky-radiometer observation network with the extension of multi-wavelength polarization measurements. *J. Remote Sens.* **2015**, *19*, 495–519.

40. Xie, Y.S.; Li, Z.Q.; Li, D.H.; Xu, H.; Li, K.T. Aerosol optical and microphysical properties of four typical sites of SONET in china based on remote sensing measurements. *Remote Sens.* **2015**, *7*, 9928–9953. [CrossRef]

41. Li, Z.; Gu, X.; Wang, L.; Li, D.; Xie, Y.; Li, K.; Dubovik, O.; Schuster, G.; Goloub, P.; Zhang, Y.; et al. Aerosol physical and chemical properties retrieved from ground-based remote sensing measurements during heavy haze days in Beijing winter. *Atmos. Chem. Phys.* **2013**, *13*, 10171–10183. [CrossRef]

42. National Centers for Environmental Prediction/National Weather Service/NOAA/U.S. Department of Commerce. 2000, updated daily. NCEP FNL Operational Model Global Tropospheric Analyses, continuing from July 1999. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. Available online: https://rda.ucar.edu/datasets/ds083.2/ (accessed on 30 July 1999).