Title

Spatiotemporal imputation of MAIAC AOD using deep learning with downscaling

Authors
Li, Lianfa
Franklin, Meredith
Girguis, Mariam
et al.

Publication Date
2020-02-01

DOI
10.1016/j.rse.2019.111584

Peer reviewed

Published in final edited form as:
Remote Sens Environ. 2020 February ; 237: . doi:10.1016/j.rse.2019.111584.

Spatiotemporal Imputation of MAIAC AOD Using Deep Learning with Downscaling

Lianfa Li[1,2], **Meredith Franklin**[1], **Mariam Girguis**[1], **Frederick Lurmann**[3], **Jun Wu**[4], **Nathan Pavlovic**[3], **Carrie Breton**[1], **Frank Gilliland**[1], **Rima Habre**[1]

[1.]Department of Preventive Medicine, University of Southern California, Los Angeles, CA, USA

[2.]State Key Laboratory of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources, Chinese Academy of Sciences, Beijing, China

[3.]Sonoma Technology, Inc., Petaluma, CA, USA

[4.]Program in Public Health, Susan and Henry Samueli College of Health Sciences, University of California, Irvine, CA, USA

Abstract

Aerosols have adverse health effects and play a significant role in the climate as well. The Multiangle Implementation of Atmospheric Correction (MAIAC) provides Aerosol Optical Depth (AOD) at high temporal (daily) and spatial (1 km) resolution, making it particularly useful to infer and characterize spatiotemporal variability of aerosols at a fine spatial scale for exposure assessment and health studies. However, clouds and conditions of high surface reflectance result in a significant proportion of missing MAIAC AOD. To fill these gaps, we present an imputation approach using deep learning with downscaling. Using a baseline autoencoder, we leverage residual connections in deep neural networks to boost learning and parameter sharing to reduce overfitting, and conduct bagging to reduce error variance in the imputations. Downscaled through a similar auto-encoder based deep residual network, Modern-Era Retrospective analysis for Research and Applications Version 2 (MERRA-2) GMI Replay Simulation (M2GMI) data were introduced to the network as an important gap-filling feature that varies in space to be used for missingness imputations. Imputing weekly MAIAC AOD from 2000 to 2016 over California, a state with considerable geographic heterogeneity, our full (non-full) residual network achieved mean R^2 = 0.94 (0.86) [RMSE = 0.007 (0.01)] in an independent test, showing considerably better performance than a regular neural network or non-linear generalized additive model (mean R^2 = 0.78–0.81; mean RMSE = 0.013–0.015). The adjusted imputed as well as combined imputed and

Authorship Contribution
Lianfa Li: Conceptualization, algorithm development, data analysis, visualization, writing & editing. **Meredith Franklin**: Critical suggestions, review and editing. **Mariam Girguis**: Data collection, review and editing. **Frederick Lurmann**: Data production, review and editing. **Jun Wu**: Review and editing. **Nathan Pavlovic**: Data collection, review and editing. **Carrie Breton**: Review and editing. **Frank Gilliland**: Review and editing. **Rima Habre**: Conceptualization, review and editing.

Competing Financial Interests
The author declares no competing financial interests.

observed MAIAC AOD showed strong correlation with Aerosol Robotic Network (AERONET) AOD (R = 0.83; $R^2 = 0.69$, RMSE = 0.04). Our results show that we can generate reliable imputations of missing AOD through a deep learning approach, having important downstream air quality modeling applications.

Keywords

aerosol optical depth; MAIAC; MERRA-2 GMI Replay Simulation; deep learning; downscaling; missingness imputation; air quality

1. Introduction

Aerosols have significant climate effects through the alteration of the radiation budget, cloud formation, atmospheric circulation, and surface temperature (Allen 2017; Kaufman et al. 2002; Stocker 2014). Ground-level aerosols are differentiated by size and are typically characterized as particulate matter (PM) with aerodynamic diameter ≤10 μm (PM_{10}) and fine particulate matter with diameter ≤2.5 μm ($PM_{2.5}$) (Hinds 1999). PM has been associated with a variety of health effects including aggravated asthma, decreased lung function, heart attacks and premature mortality (Brauer et al. 2016; EPA 2015; O’Neill et al. 2013; WHO 2013). Accurate estimation of the spatiotemporal variability of aerosols is important for understanding their role in climate change (Voiland 2010) and to reliably estimate exposures of PM_{10} and $PM_{2.5}$ for human exposure and health effects studies (Li et al. 2015). Given the limitations of sparse surface monitoring networks such as state and national PM monitoring networks, or the AErosol RObotic NETwork (AERONET) (Holben et al. 1998), researchers have begun to rely on satellite observations of AOD to characterize aerosols, particularly over large geographic areas.

Since 2000, the Moderate Resolution Imaging Spectroradiometer (MODIS) instruments on the polar orbiting TERRA and AQUA satellites have provided daily AOD retrievals with global coverage. The Dark Target (DT) algorithm and the complementary Deep Blue (DB) algorithm have been used to retrieve AOD at a 10 or 3 km resolution for diverse land cover including bright land surfaces (Levy et al. 2013). In recent years, a new algorithm, the Multiangle Implementation of Atmospheric Correction (MAIAC) algorithm (Lyapustin et al. 2011a; Lyapustin et al. 2011b) was developed to simultaneously retrieve surface bi-directional reflection function and AOD over bright and dark surfaces from MODIS at 1 km resolution. MAIAC uses time series to divide surface and atmospheric contributions and spatial information to characterize the surface bidirectional distribution function (BRDF), and in the aerosol retrieval it has improved atmospheric correction, cloud and snow detection compared to DT and DB algorithms (Lyapustin et al. 2018). As a high spatiotemporal-resolution product, MAIAC AOD was used in several studies to estimate and characterize spatiotemporal variations of aerosols and surface $PM_{2.5}$ (Di et al. 2016; Hu et al. 2017; Hu et al. 2014; Just et al. 2015; Kloog et al. 2015; Lee et al. 2011; Xiao et al. 2017; Xie et al. 2015).

In general, the retrieval of satellite aerosol products relies on clear skies and appropriate surface conditions. Cloud cover (Singh et al. 2017), snow (Bai et al. 2016) and bright surfaces (Lyapustin and Wang 2016) can result in significant proportions of missing data. A study examining the Yangtze River Delta of China found AOD were missing over >60% of the 2013–14 study period (Xiao et al. 2017). As cloud and surface conditions are geographically and temporally dependent, the missing data problem is non-random in nature (Polit and Beck 2012; Zhang and Reid 2009) making gap-filling and imputation more complex.

A variety of methods have been developed to gap-fill AOD by using spatial neighborhood information (Kloog et al. 2012), lowering the cloud screen criteria (Van Donkelaar et al. 2011) and improving retrievals over bright surfaces with the Dense Dark Vegetation algorithm (Li et al. 2012). Recently, advanced statistical methods have been developed including a two-step method combining city- and season-specific linear models plus ordinary Kriging (Lv et al. 2016), flexible generalized additive models (GAM) that allow for non-linear functional forms (Xiao et al. 2017), and feed-forward (neural) networks (Di et al. 2016). Spatial approaches such as nearest neighbors, Kriging, and GAM tend to have limited generalization (validation R^2 with AERONET AOD: 0.18–0.44) or incomplete spatial coverage of imputation (Kloog et al. 2011; Li et al. 2012; Lv et al. 2016; Van Donkelaar et al. 2011; Xiao et al. 2017). Better performing imputations often rely on external information such as outputs from chemical transport models including the Community Multi-scale Air Quality Model (CMAQ) (Di et al. 2016; Hu et al. 2017; Xiao et al. 2017) or cloud fraction (Xiao et al. 2017); however, access to these model outputs can be limited particularly for large geographic areas. Feed-forward neural networks (Di et al. 2016), when used for regression, are subject to saturation and degradation of accuracy with increased hidden layers (He et al. 2016a; Srivastava et al. 2015). To remedy this issue, residual connections of identity mapping can be introduced into the neural network to improve the learning efficiency, as shown in many applications of convolutional neural network (CNN) in deep learning (He and Sun 2015; He et al. 2016a, b). Due to the substantial amount of missing MAIAC AOD observations, it is difficult to use the CNN based deep learning methods since CNN requires complete images or images with limited random missing values for training.

We present a novel approach that incorporates deep learning to robustly impute a long time series of MAIAC AOD over a large heterogeneous region. We leverage gridded meteorological data and AOD from the Modern-Era Retrospective analysis for Research and Applications, Version 2, GMI Replay Simulation (M2GMI) (Strode et al. 2019) to provide both spatial and temporal information into an autoencoder-based deep residual network. This deep learning framework has resulted in optimal results in other air quality prediction applications (Li et al. 2018a). We impute weekly missing MAIAC AOD for 17 years (2000 to 2016) over California, evaluate our imputations against AERONET AOD, and compare them with those derived from more traditional GAM and regular neural network approaches.

2. Materials

2.1. Study Region

The study region (Fig. 1) is the State of California, which has an area of 423,970 km^2 extending from approximately −124°65’ and −114°13’ west to east longitude and 32°51’ to 42°01’ north to south latitude. California encompasses a variety of topographic, land-use and population characteristics (Fast et al. 2014) in addition to meteorological processes with significant spatiotemporal variability as a result of turbulent vertical mixing affecting dilution and chemical processes of aerosols (Aan de Brugh et al. 2012). These characteristics result in complex spatiotemporal variability of aerosols compared to central and eastern regions of the United States (US).

2.2. Data

2.2.1 MAIAC AOD—The MAIAC algorithm retrieves AOD at 1 km resolution from MODIS TERRA and AQUA satellites, with equatorial crossing at approximately 10:30 AM and 1:30 PM local time, respectively. We acquired MAIAC AOD covering California for 17 years from February 28, 2000 to December 31, 2016 from the Land Processes Distributed Active Archive Center (LP DAAC) (https://lpdaac.usgs.gov/news/release-of-modis-version-6-maiac-data-products), and extracted AOD at 550 nm with quality assurance flags and corresponding surface reflectance. Quality assurance flags indicating cloud, land, water, or snow contamination (including the adjacency mask of cloud/snow) were used to remove invalid AOD values.

2.2.2 AERONET AOD—Level 2 quality-assured AERONET AOD (version 3) was acquired for the 17-year study period from 35 sites across California (https://aeronet.gsfc.nasa.gov/) (see Fig. 1 for the spatial distribution of the sites). The 5-minute AERONET data were averaged over 60 min intervals to match the satellite overpass times, and were interpolated to 550 nm using spectral linear interpolation in the log-log space between 440 and 600 nm, the two nearest wavelengths (Eck et al. 1999; Franklin et al. 2017). AERONET AOD served as the “ground truth” to validate our imputed MAIAC AOD.

2.2.3 Meteorology—Meteorological variables were extracted from daily high resolution (~4 km, 1/24th degree) surface level meteorological data available for the contiguous US from 1979 to present (http://www.climatologylab.org/gridmet.html) (Abatzoglou 2011). Daily minimum air temperature (°C), maximum air temperature (°C), wind speed (meters/second, m/s), specific humidity (grams of vapor per kilogram of air, g/kg), daily mean downward shortwave radiation (watt/meter^2, w/m^2) and accumulated precipitation (millimeters of rain per meter^2 in 1 h, mm/m^2) were extracted and averaged to weekly values.

2.2.3 MERRA-2 Global Modeling Initiative Replay Simulation—The MERRA-2 GMI Replay Simulation (M2GMI, https://acd-ext.gsfc.nasa.gov/Projects/GEOSCCM/MERRA2GMI) is a global reanalysis data product, which, similar to its predecessor MERRA-2 (Brauer et al. 2016; Randles et al. 2017), assimilates multiple aerosol remote sensing, emissions, and meteorological data using the Goddard Earth Observing System Model (GEOS) but further incorporates aerosols, chemistry, atmosphere, land, ice, and ocean biogeochemistry. It provides 0.5° x 0.625° gridded total column aerosol optical depth and estimates of surface level sea salt, black carbon, dust, organic carbon, sulfates and $PM_{2.5}$ across our entire study period at daily time resolution. M2GMI has high temporal resolution (3 hours to daily) but coarse spatial resolution (approximately 50 km in the latitudinal direction). With these advances, it provides consistent and reliable regional estimates of aerosols for a long period of time, which is integral to our MAIAC AOD imputation. For a more detailed description of M2GMI, refer to Strode et al. (2019). For the 2000–2016 study period we acquired M2GMI total aerosol extinction AOD, which is an assimilated column and species integrated quantity that includes observations from NASA and NOAA satellites and ground-based measurements (NASA 2018; Strode et al. 2019). For simplicity, we refer to MERRA-2 GMI Replay Simulation AOD as M2GMI throughout this paper.

2.2.4 Coordinates and Elevation—The central coordinates of each MAIAC grid cell were extracted and used to capture spatial autocorrelation in our models. Elevation at 30m-resolution obtained from the GoogleMaps API was averaged over each 1 km MAIAC grid cell and used as a model variable.

3. Methods

Daily MAIAC AOD observations were preprocessed (Section 3.1) through a variety of steps to generate per-pixel weekly MAIAC AOD (Fig. 2). A deep learning modeling framework that includes two core components, a deep residual network based on an autoencoder and downscaling algorithm (Section 3.2 and 3.3) plus ensemble learning (Section 3.4) was developed based on the weekly data (Fig. 3). Model validation (Section 3.5) and prediction adjustment (Section 3.6) with AERONET data were conducted to reduce biases in the imputed AOD.

3.1. Preprocessing MAIAC AOD

There are seven steps for preprocessing MAIAC AOD images (Fig. 2): 1) bilinear resampling to re-project the Level 2 (L2) AOD to a local projection [Universal Transverse Mercator (UTM) zone 11], 2) filtering outliers and noise according to the reported range of valid AOD values (Lyapustin 2018) (i.e. remove AOD less than 0 and greater than 3), 3) applying quality assurance flags to remove observations contaminated by cloud or snow, 4) creating per-pixel daily averages of Aqua and Terra AOD, 5) fusing Aqua and Terra AODs: when both AODs were available in a pixel, their average was computed; if only one of the AODs was available, a GAM regression trained on the samples with both AODs available was applied to predict the missing Terra or Aqua AOD and then their average was computed, 6) mosaicking and cropping all MAIAC tiles for California over the study period, and 7) calculating weekly AOD averages from daily AOD preprocessed using steps 1–6. The resultant weekly MAIAC AOD are considered our analytic sample with which imputing is conducted to fill in the missing gaps.

A secondary preprocessing step was conducted to derive monthly MAIAC AOD from the daily observations after applying steps 1–6. However, based on sensitivity analyses, monthly averages were only calculated for grid cells having valid MAIAC AOD (from TERRA and/or AQUA) for at least 60% of the days of a natural month. We used the monthly averages as inputs to the deep residual network to capture spatial and longer-term temporal variability in the imputations. For locations with no valid monthly average AOD, we used the other covariates to re-train the imputation models, referred to as non-full models (Section 3.4). Correspondingly, the models trained using all the covariates including monthly averages are referred to as full models.

3.2. Autoencoder-based Deep Residual Network

The core component of the imputation framework (Fig. 3) is a deep residual network based on an autoencoder. An autoencoder (Fig. 3–c) is a neural network that has the same variables in the input and the output layers, typically one or more encoding layers, one middle coding layer, and one or more decoding layers (Kingma and Welling 2013; Liou et al. 2014). An autoencoder aims to reconstruct and recover the input variables in the output layer. Typically each encoding layer has a corresponding decoding layer with the same number of nodes (variables in a hidden layer), making it symmetrical in structure. In practice, by introducing multiple hidden layers in an encoding stage, each with decreasing numbers of nodes, the high dimensional input data are decomposed to construct powerful compact latent representations or independent principal components to reduce the input data's dimensionality (Baldi and Hornik 1989) in the middle coding layer (latent coding layer) that is beneficial for training and generalization (Jolliffe 2002). Therefore, for correlated input variables, an autoencoder resembles a principal component analysis as it reduces multicollinearity through the extraction of independent components in the latent layer (albeit through a non-linear transformation as opposed to PCA, which uses a linear transformation).

In the network developed to impute AOD, we adapt an autoencoder framework that has the same number of nodes for the encoding and decoding layers with residual connections between each shallow layer in the encoder to each deep layer in the decoder (Fig. 3–c). The mirrored/symmetrical network is a natural option to implement residual connections.

The network was trained on 1x1 km pixels that had weekly MAIAC AOD for three continuous weeks. Using this time-stratified approach to capture short-term temporal variation, an index was defined as (−1, 0, 1) for the three weeks with the middle week as the target week for prediction.

Our deep residual network contains 15 input variables (covariates): minimum and maximum temperature, wind speed, specific humidity, shortwave radiation, precipitation, M2GMI AOD, elevation, monthly average MAIAC AOD, time index, and UTM projected latitude and longitude. To account for spatial variation, the coordinates were introduced to the deep residual network as a combination of linear terms, quadratic terms, and as an interaction. For all input covariates, preprocessing was applied to filter invalid values or outliers using the quality flags and upper fences (Iglewicz and Hoaglin 1993), and each was normalized using its mean and standard deviation (Freedman et al. 2007). Comparatively, AOD was not normalized since it had a small range of valid value (0–3) with no distinct difference in the value scale from the other normalized covariates.

There are correspondingly 16 output variables: the 15 inputs plus the target variable, $\hat{y}$, unobserved MAIAC AOD (Fig. 4). The autoencoder topology consists of the encoder with 15, 128, 64, and 32 nodes for the 4 hidden layers, the middle coding layer (i.e. the latent space representation) with 16 nodes, and the decoder, which is symmetric to the encoder with 32, 64, 128, and 15 nodes for the 4 hidden layers. The network has the following loss function:

$$L(\theta_{\mathrm{W,b}}) = \frac{1}{N}\left[\ell_{\mathrm{y}}\left(\mathbf{y}, \hat{\mathbf{y}}_{\theta_{\mathrm{W,\ b}}}(\mathbf{x})\right) + \ell_{\mathrm{x}}\left(\mathbf{x}, f_{\theta_{\mathrm{W,\ b}}}(\mathbf{x})\right)\right] + \Omega\left(\theta_{\mathrm{W,b}}\right) \tag{1}$$

where N is the training sample size of each 3-week strata, $\mathbf{y}$ represents observed MAIAC AOD, $\hat{\mathbf{y}}_{\theta_{\mathrm{W,b}}}$ is the estimated (imputed) MAIAC AOD, $\theta_{\mathbf{W,b}}$ denotes the parameters of the weights, $\mathbf{W}$ and the bias, $\mathbf{b}$ of input, hidden, output layers and related batch normalization etc., $\mathbf{x}$ represents the Nx15 matrix of input covariates, $f_{\theta_{\mathrm{W,b}}}(\mathbf{x})$ represents the output matrix of $\mathbf{x}$ determined by $\theta_{\mathbf{W,b}}$, $\ell_{\mathbf{y}}$ and $\ell_{\mathbf{x}}$ denote the loss functions for the target variable $\mathbf{y}$ and the input covariates $\mathbf{x}$, respectively, and $\Omega(\theta_{\mathbf{W,b}})$ represents the regularizer for the weights and bias, $\mathbf{W}$ and $\mathbf{b}$. We use the mean square error (MSE) loss for $\ell_{\mathbf{y}}$ and $\ell_{\mathbf{x}}$ given that we are conducting regression of a continuous variable (AOD).

The introduction of $\ell_{\mathrm{x}}\left(\mathbf{x}, f_{\theta_{\mathrm{W,b}}}(\mathbf{x})\right)$ in (1) makes possible sharing of parameters between the target variable ($\mathbf{y}$, MAIAC AOD) and the covariates ($\mathbf{x}$ also as the outputs). Our training samples are large, ranging from approximately 100,000 to over 600,000 observations in each three-week grouping. The shared parameters in these training samples effectively work as a regularizer to constrain the target variable from over-fitting, as demonstrated in several other applications of deep learning (Goodfellow et al. 2016; Sun et al. 2014; Zhang et al. 2016). Sensitivity analyses show that without such sharing, generalization of the networks in prediction is reduced.

Residual connections provide shortcuts from the encoding layer to the decoding layer, boosting efficient back-propagation of errors in network learning. Assuming $\mathbf{p}_l$ is the input and $\mathbf{q}_l$ is the output for the shallow layer, l, in encoding, and $\mathbf{p}_L$ is the input and $\mathbf{q}_L$ is the output for the deep layer, L, in decoding, the residual connections for the deep layer output in decoding step are defined by:

$$\begin{aligned} q_L &= \mathbf{p}_l + f_L(\mathbf{p}_L, \mathbf{W}_L) \\ &= \mathbf{p}_l + f_L(g_L(f_l(\mathbf{p}_l, \mathbf{W}_l)), \mathbf{W}_L) \end{aligned} \tag{2}$$

where $\mathbf{W}_l$ and $\mathbf{W}_L$ represent weight parameters for the shallow and deep layer inputs, $\mathbf{p}_l$ and $\mathbf{p}_L$ respectively, f_L and f_l denote activation function for $\mathbf{p}_L$ and $\mathbf{p}_l$, respectively, and $\mathbf{p}_L = g_L(f_l(\mathbf{p}_l, \mathbf{W}_l))$ where g_L denotes the function of $\mathbf{p}_l$ for $\mathbf{p}_L$.

Based on automatic differentiation (Baydin et al. 2018), the gradient of the loss functions for the shallow layer input, $\mathbf{p}_l$ is:

$$\begin{aligned} \frac{\partial l}{\partial \mathbf{p}_l} &= \frac{\partial L}{\partial f_L(\mathbf{q}_L)} \cdot \frac{\partial f_L(\mathbf{q}_L)}{\partial \mathbf{q}_L} \cdot \frac{\partial \mathbf{q}_L}{\partial \mathbf{p}_l} \\ &= \frac{\partial L}{\partial f_L(\mathbf{q}_L)} \cdot \frac{\partial f_L(\mathbf{q}_L)}{\partial \mathbf{q}_L} \cdot \left(1 + \frac{\partial}{\partial \mathbf{p}_l} f_L(g_L(f_l(\mathbf{p}_l, \mathbf{W}_l)), \mathbf{W}_L)\right) \end{aligned} \tag{3}$$

where the constant term, 1, makes it possible to directly back-propagate errors from the deep layer, L to the shallow layer, l. Multiple residual connections nested in the internal autoencoder with deep layers together reduces vanishing of the gradient and degradation of accuracy (Li et al. 2018a).

For each hidden layer in Fig. 4, we used the rectifier activation function (ReLU) that, due to good gradient propagation and efficient computation, is widely used in many modern deep learning systems. ReLU is defined as the positive part of its argument (Hahnloser and Seung 20001):

$$Act(p) = \max(0, p) \tag{4}$$

where p is the input to a neuron and $Act\,(p)$ is the activation's output.

For the output layer, linear activation was used for regression of the output layer. Sensitivity analysis showed that this configuration of semi-linear activation functions effectively prevented the gradients from premature saturation in regression. Batch normalization was also added in each hidden layer to prevent or reduce covariance shift for effective learning (Ioffe and Szegedy 2015).

The multi-node output works as a regularizer by sharing the parameters across each output node (same as the covariates and target variable) as previously mentioned. The extra regularizer, $\Omega(\theta_{\mathbf{w,b}})$ in (1) was unnecessary in our test. The optimization of gradient descent was used to train the residual network.

We implemented the Keras version of autoencoder-based deep residual network in python and R (the library or package of resautonet).

3.3. Downscaling of MERRA-2 GMI Replay Simulation AOD with Deep Residual Network

Our imputation framework relies heavily on M2GMI AOD as an input variable, however its native spatial resolution is far coarser (50 km) than our target parameter, MAIAC AOD (1 km). In model development we found that our initial imputations had spatial variation mimicking that of the native 50 km M2GMI resolution, and that adding a prepossessing downscaling step achieved better spatial alignment of the inputs to the target 1 km spatial resolution. Thus, through a separate deep network similar to that described above, iterative downscaling was conducted to downscale 50 km M2GMI AOD to 1 km spatial resolution. Sensitivity analyses showed that the downscaling residual deep autoencoder was optimized when altered to produce a single output (1 km M2GMI AOD) rather than multiple shared outputs as described for the imputation model.

We used a reduced set of input covariates for downscaling: the projected geographic coordinates of the grid cells and 30 m elevation. Similar to in imputation, these covariates were normalized in preprocessing but M2GMI AOD was not normalized since it had a small value scale (0–2), almost consistent or just having slight differences from the normalized covariates. Given Y_k (k=1,…,R) is original M2GMI AOD in k=1,…,R grid cells at 50 km resolution, and $\hat{y}_k(k = 1, \ldots, r)$ is the M2GMI AOD estimate to be downscaled in k=1,…, r grid cells at 1 km resolution, we assume each coarse-resolution cell encapsulates n_f finely resolved cells. In initialization, the 1 km grid cells were directly assigned values from the coincident coarse resolution cell. During iteration, the estimators at the fine resolution was then adjusted in order to make the average of these adjusted 1 km estimators equal to the value of the coarse-resolution grid cell that spatially covered them:

$$\frac{1}{|F_m|}\sum_{i \in F_m} \hat{y}_i = Y_m,\ m = 1, 2,\ \ldots, R \tag{5}$$

where F_m denotes the set of fine-resolution grid cells that are spatially covered by the m^{th} coarse-resolution grid cell. Thus, we have n_f=/F_m/.

For the t^{th} iteration, we used the following formula to ensure (5):

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} \cdot \frac{Y_i}{\frac{1}{|F_m|}\sum_{j \in F_m} \hat{y}_j^{(t-1)}}, i \in F_m \tag{6}$$

At each iteration, the 1 km predictions from the previous step (t-1) were adjusted so that their average equals the 50 km M2GMI AOD (Y_k), and then with the normalized 1 km covariates (i.e., projected latitude, longitude, and elevation) the model is re-trained for the next step (t). Iterations proceeded until the mean of the absolute values of the difference in the fine-resolution cells between the last two iterations, i.e., $\frac{1}{|F|}\sum\left|\hat{y}_i^{(t)} - \hat{y}_i^{(t-1)}\right|$ (|F|, the total number of samples) satisfies a stopping criterion (SC) (i.e., >=SC) and a maximum number of iterations was also set to prevent slow convergence. The specific procedure is presented in Fig. 5. By sensitivity analysis, we selected [32, 16, 8, 4] as the optimal number of nodes for the encoding layers of the downscaling deep residual network.

3.4. Full and non-full models, bagging for ensemble predictions

As described above, monthly MAIAC AOD was not computed for grid cells with less than 60% of the daily observations over a month, so we developed two imputation models: full and non-full. The full models consisted of monthly AOD as an input, where available, in addition to meteorological and spatial input variables (15 total, described above). When monthly AOD was not available, the non-full model was identical to the full model except without monthly AOD (14 input variables total).

Random optimization of mini-batch gradient descent (Goodfellow et al. 2016) was used to learn, and to obtain stable predictions and reduce uncertainty, we used bootstrap aggregation (bagging) to train 10 multiple residual networks. Bagging reduces correlation between the models and boosts the accuracy of ensemble predictions.

Training was conducted respectively for the set of 879 weekly 1 km samples from 2/28/2000 (earliest MODIS AOD data) to 12/31/2016. In total, 879 models were trained because as described above, the time-stratified training samples consisted of data for both the target week and the week immediately before and after to ensure a sufficient sample size for training while maintaining computational tractability. Further, this approach allowed for local temporal variation in the associations between the input variables and MAIAC AOD while controlling for confounding effects caused from mixing samples at vastly different temporal scales.

3.5. Tuning of hyperparameters and validation

We used grid search to retrieve a local optimal solution for the residual neural network hyperparameters including mini-batch size, learning rate, network depth and scale, and activation functions.

For both the imputation and downscaling residual networks, the sample data was randomly split into three parts (with the stratifying factor of week index): approximately 63.2% of the samples were used for model training, one half of the remaining 36.8% samples were used for model validation (adjusting the hyperparameters in learning), and the other half was completely left out and used as an independent test. Training, validation and test R^2 and RMSE are used as statistical summaries of model performance overall, and for selected weeks for illustration.

We compare the deep residual network-based M2GMI AOD downscaling results to those obtained by a typical approach (Malone et al. 2012) that uses GAM, and the imputations to those obtained from a regular neural network with similar structure but without residual connections as well as from GAM.

Furthermore, to examine the influence of downscaling on the imputations, we performed a leave-one-county-out cross validation for the deep residual networks using 29 weeks of MAIAC AOD from 2000 to 2016 with little missing data (>90% valid weekly MAIAC AOD values). In this cross validation, data from all of the grid cells from one of California's 58 counties were removed from the training samples, and two imputation models (with original and downscaled M2GMI AOD) were trained using the samples from the remaining 57 counties. This procedure was iterated until all the cells within all the counties were predicted as the test samples. Imputed surfaces as well as mean R^2 and RMSE are reported.

To evaluate the reliability and fidelity of the developed method, we compared observed and/or imputed MAIAC AOD against AERONET AOD from the coincident MAIAC pixel, as well as for multi-pixel averages within a series of circular buffers at 1, 3, 5, 7, and 9 km around each AERONET site. Specifically, the following validations were conducted:

1. Observed MAIAC AOD (no imputation) vs. AEORNET AOD.
2. Randomly selected MAIAC AOD vs. AERONET AOD. In this validation, approximately 18.2% of the weekly observed MAIAC AOD was set aside from model training, and independent predictions were made by the trained models. Since the samples of an independent test were randomly selected at the AERONET site no buffer averages were examined.
3. Validation of imputed MAIAC AOD (no observed AOD included) vs. AERONET AOD.
4. Validation of the combined observed and imputed MAIAC AOD vs. AERONET AOD.

Test performance metrics include Pearson's correlation, R^2 and RMSE of the coincident pixel-AERONET site match and the aforementioned buffer averages where applicable.

3.6. Bias Correction with AERONET Data

Despite improved detection over regions with cloud and snow and quality screening in data preprocessing, it has been found that MAIAC AOD can be an overestimate of measured “ground truth” AOD (Li et al. 2015). To address this issue, a student’s *t*-test of paired samples was conducted to examine whether MAIAC AOD was statistically significantly greater than (i.e. overestimated) ground truth AOD at each AERONET site in the study region. If the criteria (*t*-test) shows overestimation, we conducted a systematic correction of the bias of both the original and imputed MAIAC AOD against AERONET AOD using a GAM adjusting for elevation, latitude, longitude, month and year. These covariates were selected according to their correlation with the difference between AERONET and MAIAC AOD. Given a small number of training samples for adjustment, the neural network based method was not used but rather GAM was used.

To evaluate the adjustment, the AERONET data were randomly split into 80% training and 20% independent test sets. Leave-one-site-out cross validation was also performed whereby all data from one AERONET site were iteratively removed from the training set and were compared to predicted AOD obtained from the GAM. Correlation, R^2 and RMSE between AERONET AOD and adjusted MAIAC AOD are reported in Results.

4. Results

4.1. Coverage and Summary of Satellite AOD and Covariates

The spatial availability of MAIAC AOD is visualized for the combined study period (2000–2016) and separately for summer (June till August) and winter (December till February) (Fig. 6). On average, the per-pixel MAIAC AOD missing rate is 41% over the study period, with far fewer missing in summer (21%) than winter (61%). Spatially, northern California has a higher proportion of missing observations (46%) than southern California (32%). The mean daily MAIAC AOD over the study period and state is 0.084 [standard deviation (s.d.): 0.017], with higher values in summer [0.10; s.d 0.024] than in winter (0.062; s.d. 0.018). Spatial distributions of multi-year and seasonal (summer vs. winter) observed AOD averages are presented in Fig. 7 where pixels with at least 10% completeness are shown. Descriptive statistics (mean and s.d) for overall and seasonal MAIAC AOD and the residual network input variables are presented in Table 1. Daily mean and daily standard deviations of MAIAC AOD for all pixels over the study region (California) from 2000 to 2016 are presented in Supplementary Data Fig. S1.

4.2. MERRA-2 GMI Replay Simulation AOD Downscaling

Maps showing the original (50 km) and downscaled (1 km) M2GMI AOD by GAM and deep residual network are shown in Fig. 8. Downscaling by GAM generates spatially smoother surfaces than by residual network but with loss of spatial variability compared to the original data. Two weeks (05/01/2000 to 05/07/2000 and 09/14/2015 to 09/20/2015) of typical missing MAIAC AOD patterns (low vs. high AOD; non-clustered vs. clustered) of M2GMI AOD across California are also shown (Fig 8). Correspondingly, for these weeks, downscaling via deep residual network had much better performance than GAM (test R^2: 0.94 vs. 0.66 and 0.87 vs 0.81; test RMSE: 0.05 vs. 0.14 and 0.004 vs. 0.006) (Table 2).

When downscaled M2GMI AOD are used as inputs to the residual network for MAIAC AOD imputation, validation tests show better model performance for residual network downscaling than GAM downscaling (independent test mean R^2 for downscaling: 0.89 vs. 0.78; RMSE: 0.08 vs. 0.14) (Table 2 and Supplementary Data Fig. S2). It is also notable that with residual network downscaling, the resultant imputations show realistic spatial heterogeneity, avoiding artifacts from coarse 50 km M2GMI AOD inputs (Fig. 9 c vs. Fig 9 d). In Fig. 9–11, a narrow color scale (0–0.4) rather than the full range (0–0.87) from the minimum to maximum AOD values was used to render the resultant images to highlight local details for the small AOD values. In the legend, “0.4+” indicates use of the red color for the AOD values equal to or larger than 0.4.

Full model and leave-one-county-out cross validation (CV) (R^2 and RMSE) show better MAIAC AOD imputation performance (Supplementary Data Table S1 and Fig. S3) and spatial smoothing (Supplementary Data Fig. S4) with the inclusion of downscaled versus non-downscaled M2GMI AOD (full model R^2: 0.85 vs. 0.82; full model RMSE: 0.0088 vs. 0.0093; CV R^2: 0.70 vs. 0.67; CV RMSE: 0.0082 vs. 0.0086).

4.3 Imputation

For the weeks between February 2000 and December 2016 we trained a total of 879 weekly residual networks. The autoencoder-based deep residual networks generally had better convergence and performance than regular feed-forward networks or GAM (Supplementary Data Fig. S5 for two typical weeks). The residual networks had higher mean (0.94) and range (0.84–0.99) test R^2, and lower mean (0.007) and range (0.004–0.023) test RMSE than regular networks [mean (range) R^2: 0.73 (0.08–0.99); RMSE: 0.015 (0.007–0.095)] and GAM [mean (range) R^2: 0.81 (0.57–0.93); RMSE: 0.013 (0.009–0.037)] (Table 3 and Supplementary Data Fig. S6). Overall, the residual network improved test R^2 by 21% over the regular network and by 13% over the GAM. When monthly AOD was unavailable as an input variable, the non-full residual network achieved a mean test R^2 of 0.84 (range: 0.48–0.99) and a test RMSE of 0.01 (range: 0.007–0.031); although with 10% decrease in test R^2, compared with full residual networks, the non-full residual network still performed better than full regular network and GAM. Scatter plots between observed and predicted MAIAC AOD for the full and non-full residual networks are shown in Supplementary Data Fig. S7. Observed and imputed surfaces for four sample weeks where a varying degree of MAIAC AOD observations are missing with different spatial patterns show that the deep residual network provides reliable imputations even with a large proportion (>80%) of missing data (Fig. 10).

4.4 Validation and Bias Adjustment with AERONET Data

In total over 2000–2016, 18,097 daily samples from 35 AERONET sites were available for validation and bias correction. From these daily measurements, we obtained 2,921 weekly averages for the validations of observed and/or imputed MAIAC AOD, and 737 weekly averages for the validation of independent test point MAIAC AOD.

The validation results (Table 4) show similar performance for the observed AOD, and the combined observed-imputed AOD: Pearson’s correlation: 0.67 vs. 0.69 for coincident pixel-point estimates, 0.75 vs. 0.74 for buffer radius of 9 km; RMSE of point estimates is the same (0.06); R^2 of point estimates: 0.44 vs. 0.45. The imputed AOD had lower correlations (0.60–0.66) with AERONET AOD and only a slightly higher RMSE (0.07) than the observed data. There is a small sample (*N*=122) to compare imputed AOD against AERONET AOD most likely because cases where MAIAC cannot generate a valid observation also tend to be when AERONET cannot measure AOT (i.e. cloudy skies). Situations where MAIAC is missing but AERONET is available are less common, resulting in a small sample for this validation that potentially led to uncertainty in these validation metrics.

Compared with the validations of imputed MAIAC AOD, the independent test MAIAC AOD had a larger sample size (*N*=737) and showed a higher correlation (0.81) and R^2 (0.61), and a slightly lower RMSE (0.53) with AERONET AOD for the point estimates, illustrating the reliability of our imputation approach.

Annual comparisons show an improvement in correlation over the study period (Supplementary Data Table S2) and monthly comparisons show a pronounced seasonal pattern with higher correlation and lower RMSE in summer than in winter (Table 5; Supplementary Data Fig. S8).

For those validations with averages of spatial buffers of 1–9 km, we see an Increasing positive influence on the association between MAIAC AOD and AERONET AOD with greater spatial averaging (Table 4 and 5, Supplementary Data Table S2 and Fig. S8).

Systematic over-estimation was noted in observed MAIAC AOD compared to AERONET AOD particularly at AERONET sites with high elevation. The time series of observed (un-imputed) MAIAC AOD vs. those of the measured AERONET AOD for three typical AERONET sites (Table Mountain, Goldstone and Monterey) are shown in Supplementary Data Fig. S9–a, c and e. The two sites of Table Mountain and Goldstone are typical inland sites that have different characteristics, especially surface reflectance and elevation, and irregular terrain with considerable difference between MAIAC AOD and AERONET AOD, compared with the coastline AERONET sites such as Monterey (Loría-Salazar et al. 2016). Further, student's *t*-tests of the samples from all the 35 AERONET sites showed that on average MAIAC AOD overestimated AERONET AOD by 0.005 (*t*=4.8, *p*-value<0.05). The difference between MAIAC and AERONET AOD was strongly correlated with elevation (Pearson's correlation: 0.72), and moderately correlated with latitude (−0.22) and longitude (0.17). After imputation, the GAM regression between MAIAC (including observed and imputed values) and AERONET AOD including elevation, latitude and longitude obtained an R^2 of 0.75 and an RMSE of 0.04.

Using leave-one-site-out cross validation, the final MAIAC AOD adjusted by GAM obtained a better correlation (0.83), R^2 (0.69) and RMSE (0.04) with AERONET AOD than un-adjusted MAIAC AOD (Supplementary Data Fig. S9–b, d and f for the times series of adjusted MAIAC AOD for Table Mountain, Goldstone and Monterey, respectively; Fig. S10 for the scatter and residual plots of the samples of all the AERONET sites). Surfaces of the adjusted MAIAC AOD for the same four weeks as shown above (Fig. 10) are presented in Fig. 11.

5. Discussion

We developed a deep learning approach to improve imputation of massive non-randomly missing MAIAC AOD over a large and heterogeneous geographic region. As a powerful tool often used in image processing (Goodfellow et al. 2016), feature extraction and prediction, convolutional neural networks cannot be directly applied for missing data imputation. As a viable alternative, we adopt an autoencoder-based residual network in a regression framework with residual connections that boost back-propagation of the errors from the deep to shallow layers. With multiple outputs in the network topology, the parameters were shared across the inputs and target output variable, which effectively prevented the models from over-fitting. By grid search for hyperparameters in deep learning including mini batch size, learning rate, number of hidden layers and numbers of nodes, we obtained a locally optimal solution that sufficiently improved effectiveness in learning.

We apply this deep residual network framework as the core component in both imputation and spatial downscaling of a key input variable. With limited available inputs (meteorology, M2GMI AOD, monthly MAIAC AOD, spatial coordinates and elevation), our approach achieved cutting-edge performance with mean test R^2 = 0.94 (range 0.85–0.99) for the imputation of a long time series of MAIAC AOD with a significant proportion of missing values (41%). Spatiotemporally varying predictor variables such as meteorology, monthly AOD and M2GMI AOD were used to capture variability of MAIAC AOD. Compared with a global model, the weekly local models trained by different values of the spatially and/or temporally varying predictor variables better captured the local temporal variability in the association between MAIAC AOD and the predictors, and reduced the confounding effects of mixing samples at vastly different temporal scales. Similar localized modeling methods were used to impute satellite AOD (Xiao et al. 2017) and estimate $PM_{2.5}$ (Li et al. 2018b). Compared to a regular neural network and GAM, the deep residual network performed 13–21% better. Xiao et al. (2017) used a GAM to impute MAIAC AOD over the Yangtze River Delta of China, which achieved an average R^2 of 0.77 (ranging from 0.48 to 0.97 in model fitting) and an R^2 of 0.44 in validation with AERONET AOD. In addition, they included more covariates than what we used in this paper, such as cloud fraction, normalized difference vegetation index and CMAQ simulations. These measures are not always publicly and readily available for the long time series imputation in certain regions. Other studies (Di et al. 2016; Just et al. 2015; Kloog 2016; Lv et al. 2016) using a variety of modeling approaches had lower performance results or lower imputation rates for missing satellite AOD. To our knowledge, this is the first study to employ advanced deep learning techniques for robust downscaling and imputation of massive missing satellite AOD.

To fuse input variables (e.g., M2GMI and meteorology) at multiple scales, simple methods such as bilinear or nearest neighbor resampling are traditionally used. Such traditional methods can introduce grid effects (bias) at coarse resolution in scenarios where the spatial resolutions of the layers are very different (Alparone et al. 2015; Baboo and Devi 2010; Wald 2002). Kriging based area-to-point prediction methods (Goovaerts 2010; Gotway and Young 2002; Pardo-Iguzquiza et al. 2011) and GAM (Malone et al. 2012) can also be used, but given a large training sample size, generalization is limited in comparison with deep learning (Goodfellow et al. 2016). Comparatively, our downscaling algorithm, that also leveraged a deep residual network, captured local heterogeneity at fine spatial scales. Our findings highlight that better downscaling performance can be achieved using deep residual networks compared to more traditional methods such as GAM.

With an internal cloud mask and snow detection, MAIAC AOD provides quality assurance metrics to enable detection and removal of invalid or highly uncertain values of AOD. Despite these quality assurance metrics, investigators have found that satellite AOD is subject to occasional overestimations when compared to AERONET AOD in locations along cloud and snow edges (Emili et al. 2011; Li et al. 2015). We observed and addressed similar artifacts, particularly in locations of high elevation. California's diverse topography, elevation, emission sources (e.g. traffic, dust, photochemical reactions) and heterogenous meteorology result in high spatiotemporal variability of aerosols. Furthermore, cloud, snow or high surface reflectance is a leading cause for missing AOD values. California has more rain, clouds and snow in winter than in summer, and it also has large desert-like areas or deserts, which results in both a greater proportion of missing values and an over-estimation of AOD (Li et al. 2015). We observe 61% missing observations in December-February in addition to lower correlation and overestimation between MAIAC AOD and AERONET AOD in winter. Despite this, the extensive validations of imputed MAIAC AOD vs. AERONET AOD demonstrated the reliability of the proposed approach. Our investigation also demonstrates that AOD overestimation can be explained and corrected by elevation, geography, and temporal trends, resulting in improved correlation between MAIAC and AERONET AOD from 0.70 to 0.83.

Widespread wildfires (Wikipedia 2018) in California occasionally caused very high AOD, e.g., in 2008, 2015 and 2016. For example, the week from 09/14/2015 to 09/20/2015, a typical wildfire season week, had a distribution of MAIAC AOD that covered high AOD values (range: 0 – 2), compared with that (range: 0 – 0.3) of a non-wildfire week from 05/01/2000 to 05/07/2000 (Supplementary Data Fig. S7). Correspondingly, AERONET AOD also presented high values (>0.6) for wildfire weeks, e.g., 06/23/2008–06/29/2008, 07/07/2008–07/13/2008, 09/07/2015–09/13/2015, 07/25/2016–07/31/2016 and 08/15/2016–08/21/2016 (Supplementary Data Fig. S10). As shown in the tests, the proposed approach reliably captured the variance of high AOD even though these samples were limitedly or sparsely distributed.

A long time series of spatially resolved MAIAC AOD with missing observation imputations having performance statistics that surpass those of previous studies provides the basis for a more complete and accurate analysis of the spatiotemporal variability of aerosols over California. Furthermore, these AODs can be leveraged to make better characterizations and inferences about ground-level $PM_{2.5}$. Without complete MAIAC AOD we would have to rely on the other external information to estimate $PM_{2.5}$ concentrations, which could result increased bias and measurement error (Paciorek and Liu 2009). The downstream impact of these errors can lead to underpowered, null or erroneous evaluation of $PM_{2.5}$ related health effects.

Our approach has important implications for reduction of estimation errors of $PM_{2.5}$ and subsequently bias of exposure estimation in evaluation of its health effects. Our approach is also easily generalized to finer spatial scale (e.g. 1 km) and other regions. With the robust imputation of MAIAC AOD from 2000 to 2016 in California, future work is to use them to make high-resolution spatiotemporal estimates of $PM_{2.5}$ across California.

6. Conclusion

Aerosol observations obtained from remote sensing suffer from limited coverage due to high surface reflectance and clouds, and the patterns of missing data are not random. We developed a robust approach for imputing missing MAIAC AOD by leveraging a deep residual neural network based on autoencoder and the MERRA-2 GMI Replay Simulation data product. Downscaling key input variables to the resolution of the target variable vastly improved the spatial representativeness of the imputations, and including monthly averaged MAIAC AOD improved model performance. For the case study of weekly MAIAC AOD imputation over California for 17 years (2000–2016), our approach achieved considerably better mean test R^2 (full model: 0.94; non-full model: 0.86) with lower test RMSE (0.007–0.01) than regular neural networks and GAMs, and our performance metrics surpassed those previously published imputation studies. In validation against data from 35 AERONET sites, the extensive tests provided a strong support to the generalization of the proposed approach, and the final adjusted MAIAC AOD had strong total correlation (0.83), illustrating the reliability of our imputations. Our approach can be easily generalized to finer temporal resolution (daily) and other regions and could be applied to different satellite data products such as aerosol observations from other instruments or observations of surface properties. Importantly, this study has downstream benefits in terms of improved exposure estimation of $PM_{2.5}$ with reduced measurement error that will ensure more reliable evaluations of $PM_{2.5}$ –related health effects.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The study was supported by the Lifecourse Approach to Developmental Repercussions of Environmental Agents on Metabolic and Respiratory Health NIH ECHO grants (4UH3OD023287) and the Southern California Environmental Health Sciences Center (National Institute of Environmental Health Sciences' grant, P30ES007048). The authors gratefully acknowledge Dr. Luke David Oman, Research Physical Scientist in the Atmospheric Chemistry and Dynamics Laboratory at NASA Goddard Space Flight Center, for providing the MERRA-2 GMI Replay Simulation data and guidance on its use in our modeling efforts. The authors also gratefully thank Dr. Melanie Follette-Cook, Dr. Pawan Gupta and Dr. Bryan Duncan for training, guidance and insights on NASA data products and their application in health and air quality studies. The MERRA-2 GMI Replay simulation is supported by the NASA Modeling, Analysis and Prediction (MAP) program and the high-performance computing resources were provided by the NASA Center for Climate Simulation (NCCS). Finally, the authors gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPUs used in this research.

REFERENCES

Aan de Brugh JMJ, Henzing JS, Schaap M, Morgan WT, van Heerwaarden CC, Weijers EP, Coe H, & Krol MC (2012). Modelling the partitioning of ammonium nitrate in the convective boundary layer. Atmos. Chem. Phys, 12, 3005–3023

Abatzoglou TJ (2011). Development of gridded surface meteorological datafor ecological applications and modelling. International Journal of Climatology, 2011

Allen B (2017). Atmospheric Aerosols: What Are They, and Why Are They So Important? In

Alparone L, Aiazzi B, Baronti S, & Garzelli A (2015). Remote Sensing Image of Fusion Boca Raton, Fl: Taylor & Francis

Baboo SS, & Devi MR (2010). An analysis of different resampling methods in Coimbatore, district. Global Journal of Computer Science and Technology

Bai Y, Wu L, Qin K, Zhang Y, Shen Y, & Zhou Y (2016). A Geographically and TemporallyWeighted Regression Model for Ground-Level PM2.5 Estimation from Satellite-Derived 500 m Resolution AOD. Remote Sensing, 8, 262

Baldi P, & Hornik K (1989). Neural Networks and Principal Component Analysis - Learning from Examples without Local Minima. Neural Networks, 2, 53–58

Baydin GA, Pearlmutter B, Radul AA, & Siskind J (2018). Automatic differentiation in machine learning: a survey. Journal of Machine Learning Research, 18, 1–43

Brauer M, Freedman G, Frostad J, van Donkelaar A, Martin RV, Dentener F, van Dingenen R, Estep K, Amini H, Apte JS, Balakrishnan K, Barregard L, Broday D, Feigin V, Ghosh S, Hopke PK, Knibbs LD, Kokubo Y, Liu Y, Ma SF, Morawska L, Sangrador JLT, Shaddick G, Anderson HR, Vos T, Forouzanfar MH, Burnett RT, & Cohen A (2016). Ambient Air Pollution Exposure Estimation for the Global Burden of Disease 2013. Environ Sci Technol, 50, 79–88 [PubMed: 26595236]

Di Q, Kloog I, Koutrakis P, Lyapustin A, Wang Y, & Schwartz J (2016). Assessing PM2.5 Exposures with High Spatiotemporal Resolution across the Continental United States. Environ Sci Technol, 50, 4712–4721 [PubMed: 27023334]

Eck TF, Holben BN, Reid JS, Dubovik O, Smirnov A, O'Neill NT, Slutsker I, & Kinne S (1999). Wavelength dependence of the optical depth of biomass burning, urban, and desert dust aerosols. Journal of Geophysical Research-Atmospheres, 104, 31333–31349

Emili E, Lyapustin A, Wang Y, Popp C, Korkin S, Zebisch M, Wunderle S, & Petitta M (2011). High spatial resolution aerosol retrieval with MAIAC: Application to mountain regions. Journal of Geophysical Research, 16, 1–12

EPA (2015). Particulate Matter Emissions In

Fast JD, Allan J, Bahreini R, Craven J, Emmons L, Ferrare R, Hayes PL, Hodzic A, Holloway J, Hostetler C, Jimenez JL, Jonsson H, Liu S, Liu Y, Metcalf A, Middlebrook A, Nowak J, Pekour M, Perring A, Russell L, Sedlacek A, Seinfeld J, Setyan A, Shilling J, Shrivastava M, Springston S, Song C, Subramanian R, Taylor JW, Vinoj V, Yang Q, Zaveri RA, & Zhang Q (2014). Modeling regional aerosol and aerosol precursor variability over California and its sensitivity to emissions and long-range transport during the 2010 CalNex and CARES campaigns. Atmospheric Chemistry and Physics, 14, 10013–10060

Franklin M, Kalashnikova OV, & Garay MJ (2017). Size-resolved particulate matter concentrations derived from 4.4 km-resolution size-fractionated Multi-angle Imaging SpectroRadiometer (MISR) aerosol optical depth over Southern California. Remote Sensing of Environment, 196, 312–323

Freedman D, Pisani R, & Purves R (2007). Statistics: Fourth International Student Edition W.W. Norton & Company

Goodfellow I, Bengio Y, & Courville A (2016). Deep Learning MIT Press

Goovaerts P (2010). Combining areal and point data in geostatistical interpolation: applications to soil science and medical geography. Mathematical Geosciences, 42, 535–554 [PubMed: 21132098]

Gotway CA, & Young LJ (2002). Combining incompatible spatial data. Journal of the American Statistical Association, 97, 632–648

Hahnloser R, & Seung SH (20001). Permitted and Forbidden Sets in Symmetric Threshold-Linear Networks. In, NIPS 2001

He K, & Sun J (2015). Convolutional neural networks at constrained time cost. In, CVPR

He KM, Zhang XY, Ren SQ, & Sun J (2016a). Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (Cvpr), 770–778

He KM, Zhang XY, Ren SQ, & Sun J (2016b). Identity Mappings in Deep Residual Networks. Computer Vision - Eccv 2016, Pt Iv, 9908, 630–645

Hinds CW (1999). Measurement of Airborne Particles, 2nd New York: John Wiley & Sons Inc.

Holben BN, Eck TF, Slutsker I, Tanre D, Buis JP, Setzer A, Vermote E, Reagan JA, Kaufman YJ, Nakajima T, Lavenu F, Jankowiak I, & Smirnov A (1998). AERONET - A federated instrument network and data archive for aerosol characterization. Remote Sensing of Environment, 66, 1–16

Hu X, Belle JH, Meng X, Wildani A, Waller LA, Strickland MJ, & Liu Y (2017). Estimating PM2.5 Concentrations in the Conterminous United States Using the Random Forest Approach. Environ Sci Technol, 51, 6936–6944 [PubMed: 28534414]

Hu XF, Waller LA, Lyapustin A, Wang YJ, Al-Hamdan MZ, Crosson WL, Estes MG, Estes SM, Quattrochi DA, Puttaswamy SJ, & Liu Y (2014). Estimating ground-level PM2.5 concentrations in the Southeastern United States using MAIAC AOD retrievals and a two-stage model. Remote Sensing of Environment, 140, 220–232

Iglewicz B, & Hoaglin CD (1993). How to detect and handle outliers In Mykytka FE (Ed.), The ASQ Basic References in Quality Control: Statistical Techniques Milwaukee: American Society for Quality

Ioffe S, & Szegedy C (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In, ICML

Jolliffe TI (2002). Principal Component Analysis (second edition) New York: Springer-Verlag

Just AC, Wright RO, Schwartz J, Coull BA, Baccarelli AA, Tellez-Rojo MM, Moody E, Wang Y, Lyapustin A, & Kloog I (2015). Using high-resolution satellite aerosol optical depth to estimate daily PM2. 5 geographical distribution in Mexico City. Environmental Science and Technology, 49, 8576–8584 [PubMed: 26061488]

Kaufman YJ, Tanre D, & Boucher O (2002). A satellite view of aerosols in climate systems. Nature, 419, 215–223 [PubMed: 12226676]

Kingma PD, & Welling M (2013). Auto-Encoding Variational Bayes. In: arXiv

Kloog I (2016). Fine particulate matter (PM2.5) association with peripheral artery disease admissions in northeastern United States Int. J. Environ. Health Res, 26, 572–577

Kloog I, Koutrakis P, Coull BA, Lee HJ, & Schwartz J (2011). Assessing temporally and spatially resolved PM2.5 exposures for epidemiological studies using satellite aerosol optical depth measurements. Atmospheric Environment, 45, 6267–6275

Kloog I, Nordio F, Coull BA, & Schwartz J (2012). Incorporating Local Land Use Regression And Satellite Aerosol Optical Depth In A Hybrid Model Of Spatiotemporal PM2.5 Exposures In The Mid-Atlantic States. Environ Sci Technol, 46, 11913–11921 [PubMed: 23013112]

Kloog I, Sorek-Hamer M, Lyapustin A, Coull B, Wang Y, Just AC, Schwartz J, & Broday DM (2015). Estimating daily PM2.5 and PM10 across the complex geo-climate region of Israel using maiac satellite-based AOD data. Atmos Environ, 122, 409–416

Lee HJ, Liu Y, Coull BA, Schwartz J, & Koutrakis P (2011). A novel calibration approach of MODIS AOD data to predict PM2.5 concentrations. Atmos. Chem. Phys, 11, 7991–8002

Levy RC, Mattoo S, Munchak LA, Remer LA, Sayer AM, Patadia F, & Hsu NC (2013). The Collection 6 MODIS aerosol products over land and ocean. Atmospheric Measurement Techniques, 6, 2989–3034

Li J, Carlson EB, & Lacis AA (2015). How well do satellite AOD observations represent the spatial and temporal variability of PM2.5 concentration for the United States? Atmos Environ, 102, 260–273

Li L, Fang Y, Wu J, & Wang J (2018a). Autoencoder Based Residual Deep Networks for Robust Regression Prediction and Spatiotemporal Estimation. In, arXiv e-prints

Li L, Zhang J, Meng X, Fang Y, Ge Y, Wang J, Wang C, Wu J, & Kan H (2018b). Estimation of PM2. 5 concentrations at a high spatiotemporal resolution using constrained mixed-effect bagging models with MAIAC aerosol optical depth. Remote Sensing of Environment, 217, 573–586

Li SS, Chen LF, Tao JH, Han D, Wang ZT, Su L, Fan M, & Yu C (2012). Retrieval of aerosol optical depth over bright targets in the urban areas of North China during winter. Science China-Earth Sciences, 55, 1545–1553

Liou CY, Cheng WC, Liou JW, & Liou DR (2014). Autoencoder for words. Neurocomputing, 139, 84–96

Loría-Salazar SM, Holmes HA, Arnott WP, Barnard JC, & Moosmüller H (2016). Evaluation of MODIS columnar aerosol retrievals using AERONET in semi-arid Nevada and California, USA, during the summer of 2012. Atmospheric Environment, 144, 345–360

Lv B, Hu Y, Chang HH, Russell AG, & Bai Y (2016). Improving the accuracy of daily PM2.5 distributions derived from the fusion of ground-level measurements with aerosol optical depth observations, a case study in north China. Environ Sci Technol, 50, 4752–4759 [PubMed: 27043852]

Lyapustin A (2018). MCD19A2 V006. In: USGS

Lyapustin A, Martonchik J, Wang YJ, Laszlo I, & Korkin S (2011a). Multiangle implementation of atmospheric correction (MAIAC): 1. Radiative transfer basis and look-up tables. Journal of Geophysical Research-Atmospheres, 116

Lyapustin A, & Wang Y (2016). MODIS Multi-Angle Implementation of Atmospheric Correction (MAIAC) Data User's Guide In

Lyapustin A, Wang Y, Laszlo I, Kahn R, Korkin S, Remer L, Levy R, & Reid JS (2011b). Multiangle implementation of atmospheric correction (MAIAC): 2. Aerosol algorithm. Journal of Geophysical Research-Atmospheres, 116

Lyapustin A, Wang YJ, Korkin S, & Huang D (2018). MODIS Collection 6 MAIAC algorithm. Atmospheric Measurement Techniques, 11, 5741–5765

Malone PB, McBratney BA, Minasny B, & Wheeler I (2012). A general method for downscaling earth resource information. Computers & Geosciences, 41, 119–125

NASA (2018). MERRA-2 GMI In

O'Neill SM, Lahm PW, Fitch MJ, & Broughton M (2013). Summary and analysis of approaches linking visual range, PM2.5 concentrations, and air quality health impact indices for wildfires. J Air Waste Manag Assoc, 63, 1083–1090 [PubMed: 24151683]

Paciorek CJ, & Liu Y (2009). Limitations of Remotely Sensed Aerosol as a Spatial Proxy for Fine Particulate Matter. Environmental Health Perspectives, 117, 904–909 [PubMed: 19590681]

Pardo-Iguzquiza E, Rodriguez-Galiano VF, Chica-Olmo M, & Atkinson PM (2011). Image fusion by spatially adaptive filtering using downscaling cokriging. ISPRS Journal of Photogrammetry and Remote Sensing, 66, 337–346

Polit DF, & Beck CT (2012). Nursing Research: Generating and Assessing Evidence for Nursing Practice (9th ed.) Philadelphia, USA: Wolters Klower Health, Lippincott Williams & Wilkins

Randles CA, da Silva AM, Buchard V, Colarco PR, Darmenov A, Govindaraju R, Smirnov A, Holben B, Ferrare R, Hair J, Shinozuka Y, & Flynn CJ (2017). The MERRA-2 Aerosol Reanalysis, 1980 Onward. Part I: System Description and Data Assimilation Evaluation. Journal of Climate, 30, 6823–6850 [PubMed: 29576684]

Singh MK, Venkatachalam P, & Gautam R (2017). Geostatistical Methods for Filling Gaps in Level-3 Monthly-Mean Aerosol Optical Depth Data from Multi-Angle Imaging SpectroRadiometer. Aerosol and Air Quality Research, 17, 1963–1974

Srivastava KR, Greff K, & Schmidhuber J (2015). Highway networks. In: arXiv:1505.00387

Stocker TF (2014). Climate Change 2013: the Physical Science Basis. Summary for Policymakers In

Strode SA, Ziemke JR, Oman LD, Lamsal LN, Olsen MA, & Liu JH (2019). Global changes in the diurnal cycle of surface ozone. Atmospheric Environment, 199, 323–333

Sun Y, Chen YH, Wang XG, & Tang XO (2014). Deep Learning Face Representation by Joint Identification-Verification. Advances in Neural Information Processing Systems 27 (Nips 2014), 27

Van Donkelaar A, Martin RV, Levy RC, da Silva AM, Krzyzanowski M, Chubarova NE, Semutnikova E, & Cohen AJ (2011). Satellite-based estimates of ground-level fine particulate matter during extreme events: a case study of the Moscow fires in 2010. Atmos Environ, 45, 6225–6232

Voiland A (2010). Aerosols: Tiny Particles, Big Impact. In: NASA

Wald L (2002). Data Fusion, Definition and Architectures--Fusion of Images of Different Spatial Resolutions Paris: Les Presses de

WHO (2013). Review of evidence on health aspects of air pollution—REVIHAAP project: Final technical report In. Bonn, Switzerland: The WHO European Centre for Environment and Health

Wikipedia (2018). List of California wildfires In

Xiao Q, Wang Y, Chang HH, Meng X, Geng G, Lyapustin A, & Liu Y (2017). Full-coverage high-resolution daily PM2.5 estimation using MAIAC AOD in the Yangtze River Delta of China. Remote Sensing of Environment, 199, 437–446

Xie Y, Wang Y, Zhang K, Dong W, Lv B, & Bai Y (2015). Daily Estimation of Ground-Level PM2.5 Concentrations over Beijing Using 3 km Resolution MODIS AOD. Environ Sci Technol, 49, 12280–12288 [PubMed: 26310776]

Zhang JL, & Reid JS (2009). An analysis of clear sky and contextual biases using an operational over ocean MODIS aerosol product. Geophysical Research Letters, 36

Zhang ZP, Luo P, Loy CC, & Tang XO (2016). Learning Deep Representation for Face Alignment with Auxiliary Attributes. Ieee Transactions on Pattern Analysis and Machine Intelligence, 38, 918–930 [PubMed: 27046839]

Highlights

- Massive non-random missingness limits satellite AOD applications.
- Residual learning of deep network boosts training efficiency for AOD imputation.
- Residual network reliably downscales coarse-scale reanalysis data.
- Adjusted satellite AOD using elevation and coordinates better correlates with AERONET AOD.
- Residual network generalizable to impute missing satellite or environmental data.

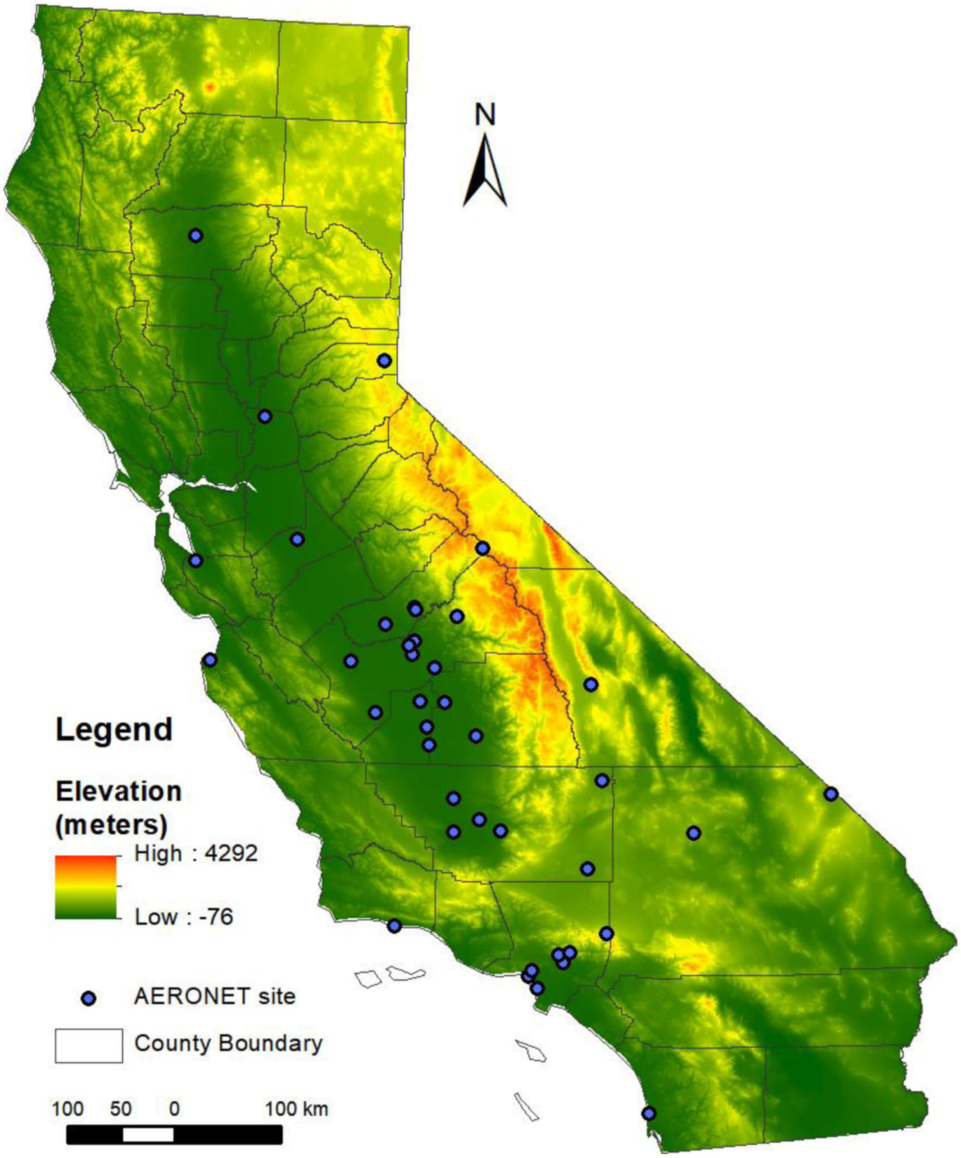


Fig. 1.

California study region showing elevation (30 m resolution) and locations of the AERONET sites.

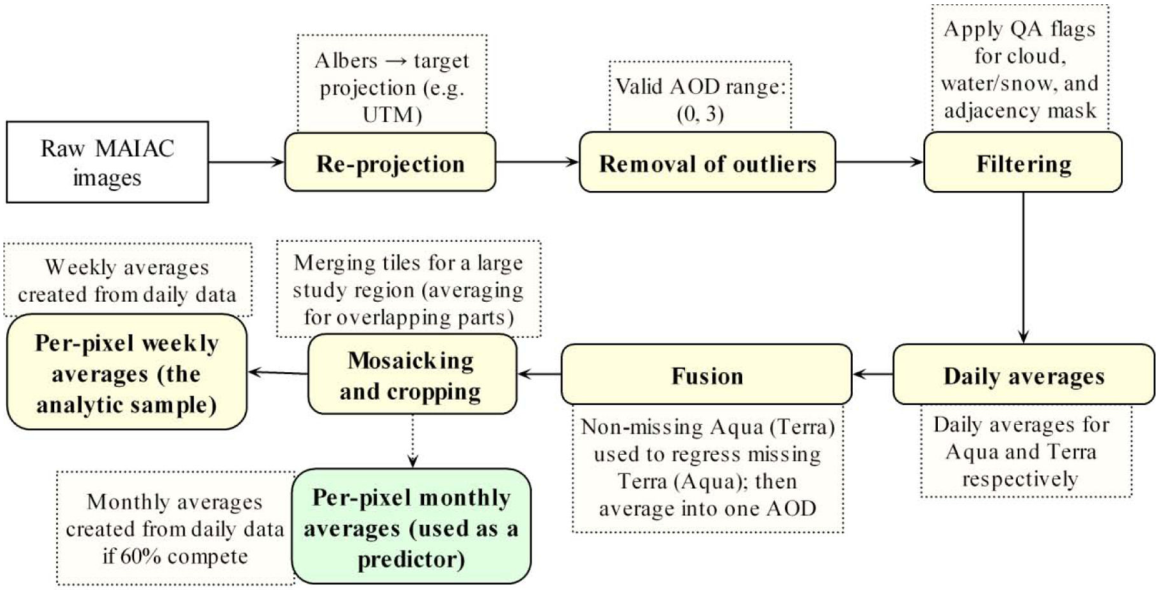

Raw MAIAC images
Re-projection
Albers → target projection (e.g. UTM)
Removal of outliers
Valid AOD range: (0, 3)
Filtering
Apply QA flags for cloud, water/snow, and adjacency mask
Daily averages
Daily averages for Aqua and Terra respectively
Fusion
Non-missing Aqua (Terra) used to regress missing Terra (Aqua); then average into one AOD
Mosaicking and cropping
Merging tiles for a large study region (averaging for overlapping parts)
Per-pixel weekly averages (the analytic sample)
Weekly averages created from daily data
Per-pixel monthly averages (used as a predictor)
Monthly averages created from daily data if 60% compete

Fig. 2.

Preprocessing pipeline of MAIAC AOD

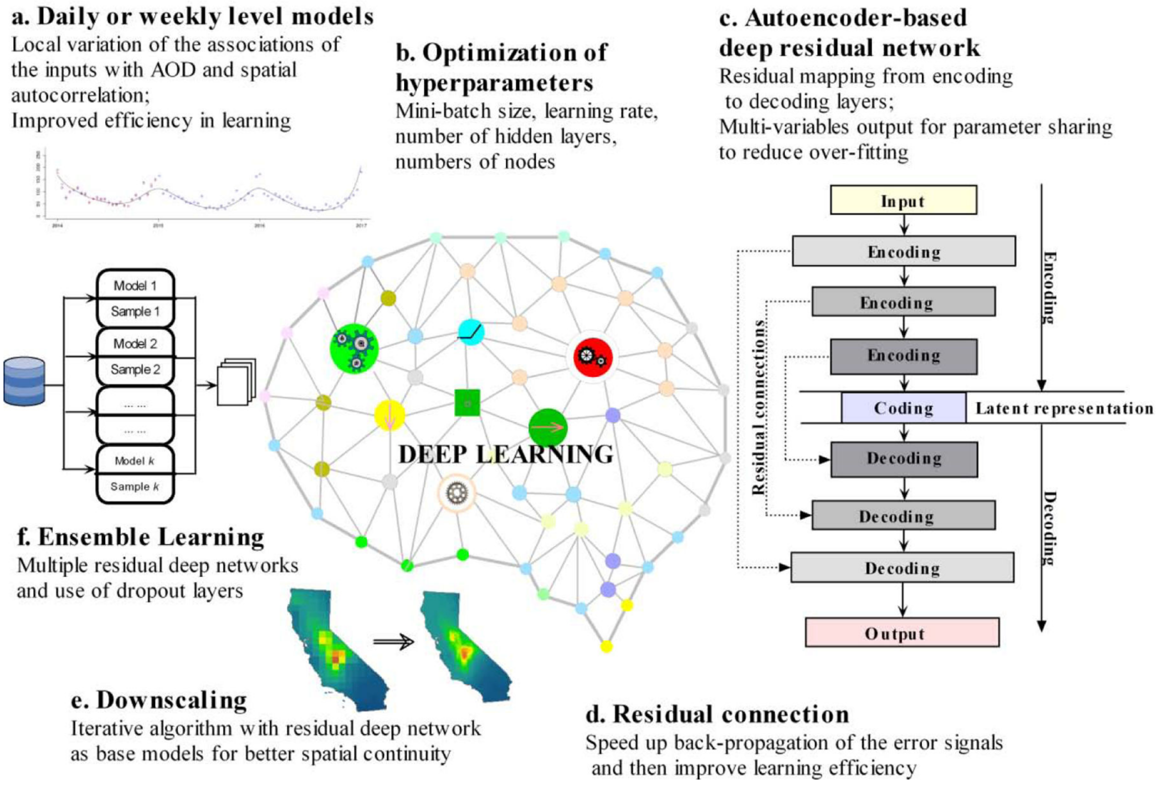

a. Daily or weekly level models
Local variation of the associations of the inputs with AOD and spatial autocorrelation;
Improved efficiency in learning
2014
2015
2016
2017
b. Optimization of hyperparameters
Mini-batch size, learning rate, number of hidden layers, numbers of nodes
c. Autoencoder-based deep residual network
Residual mapping from encoding to decoding layers;
Multi-variables output for parameter sharing to reduce over-fitting
Input
Encoding
Encoding
Encoding
Coding
Latent representation
Decoding
Decoding
Decoding
Output
Encoding
Decoding
Residual connections
Model 1
Sample 1
Model 2
Sample 2
... ...
... ...
Model k
Sample k
DEEP LEARNING
f. Ensemble Learning
Multiple residual deep networks and use of dropout layers
e. Downscaling
Iterative algorithm with residual deep network as base models for better spatial continuity
d. Residual connection
Speed up back-propagation of the error signals and then improve learning efficiency

Fig. 3.

Autoencoder-based deep residual neural network modeling framework

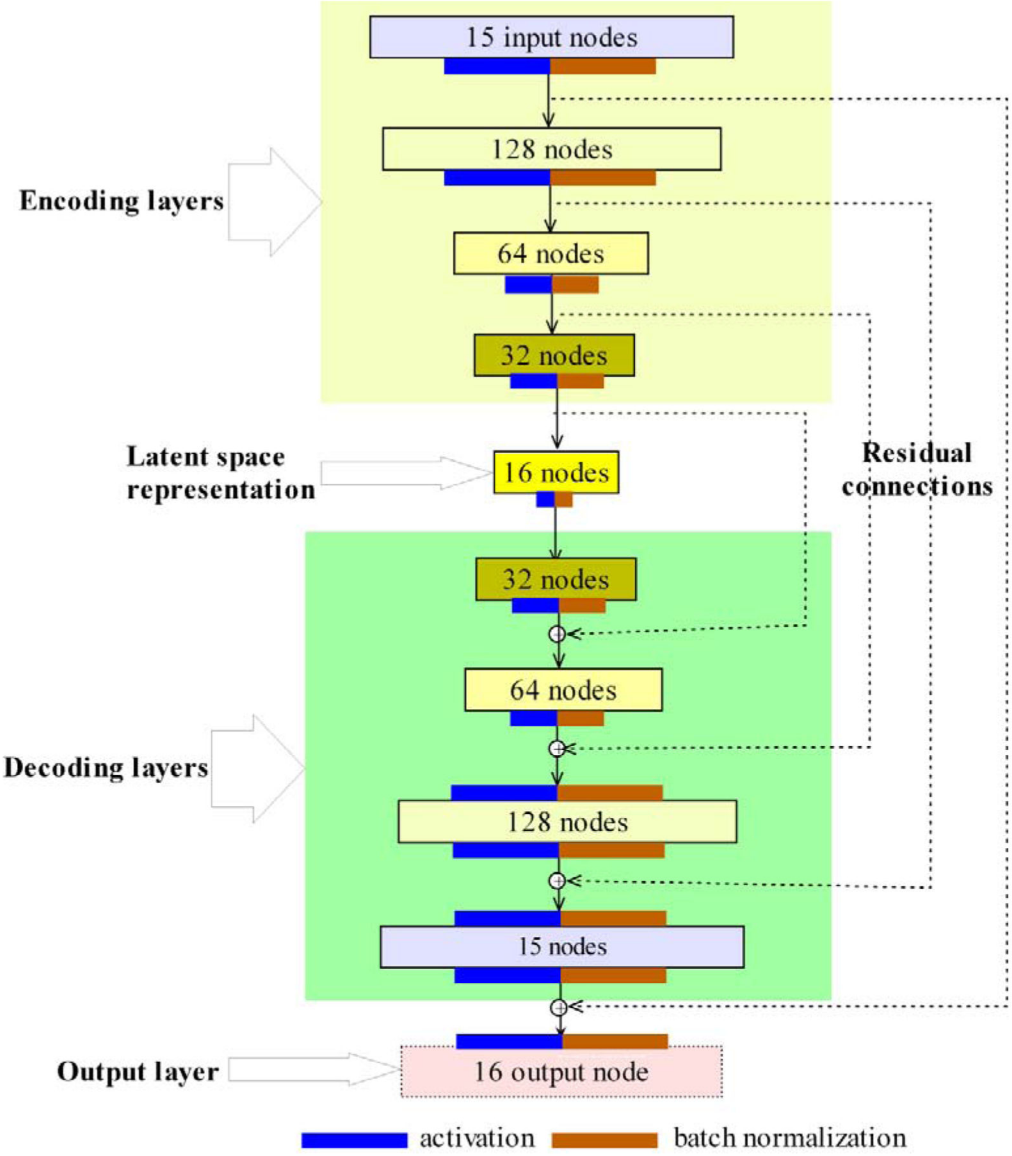
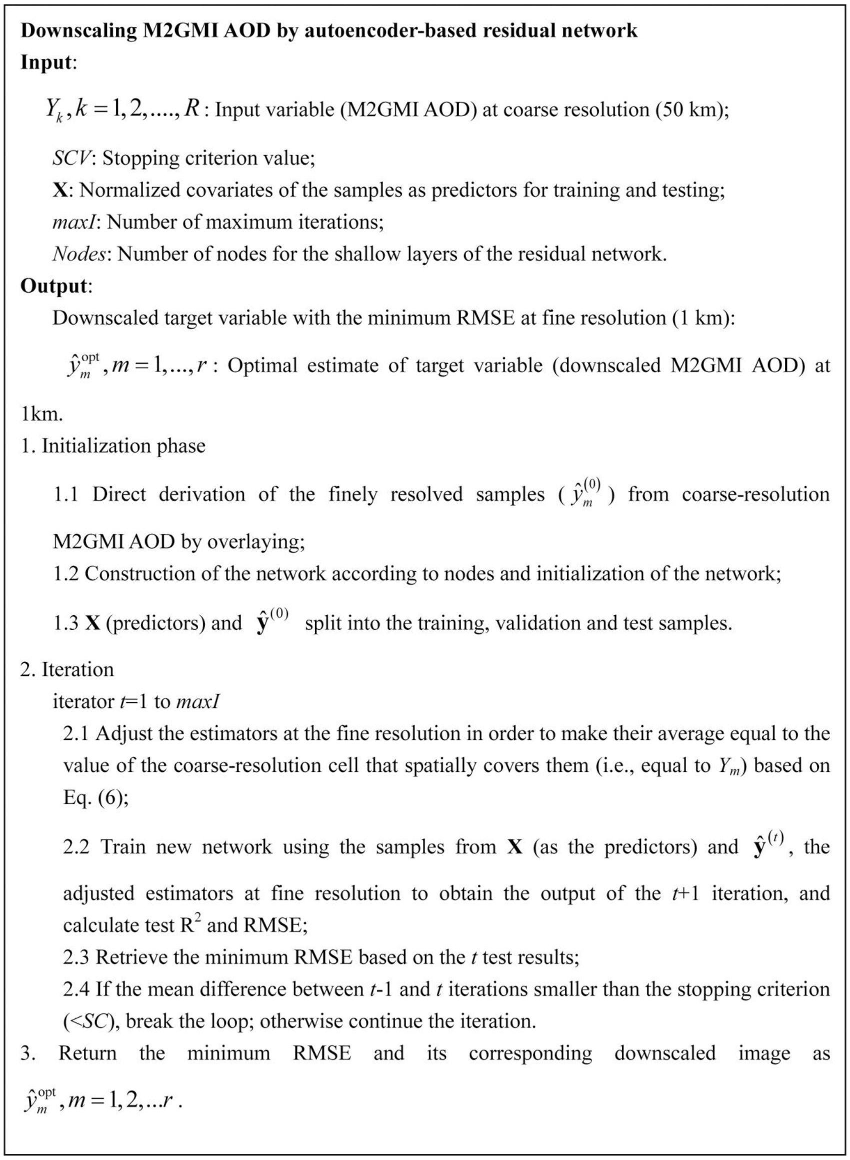


Fig. 4.

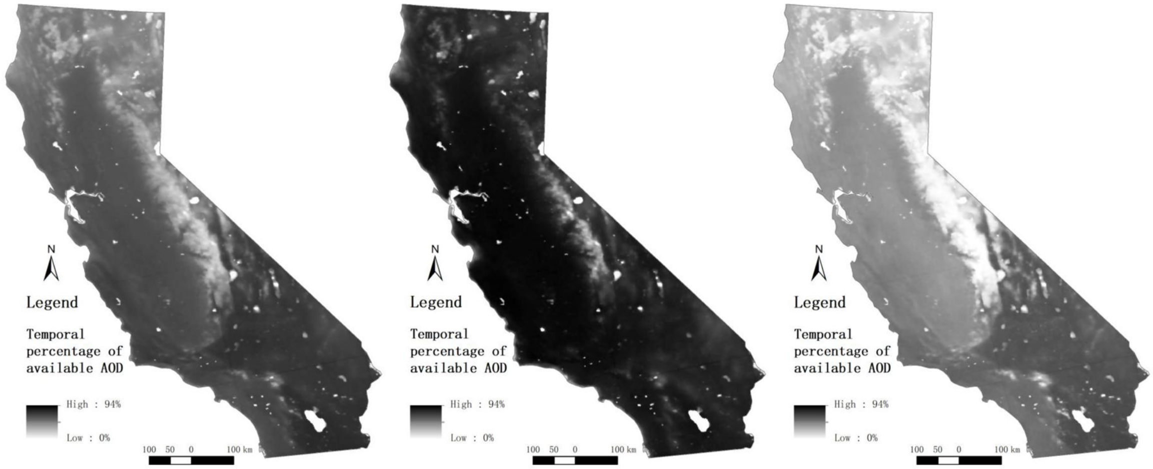
Autoencoder-based deep residual network setup. Dashed lines form the residual connections of the 4 hidden layers from the encoder (yellow box) to the decoder (green box); the output at every layer is activated (blue rectangles) and batch normalized (brown rectangles).



```
Downscaling M2GMI AOD by autoencoder-based residual network
Input:
    $Y_k, k = 1, 2, ...., R$ : Input variable (M2GMI AOD) at coarse resolution (50 km);
    $SCV$: Stopping criterion value;
    $\mathbf{X}$: Normalized covariates of the samples as predictors for training and testing;
    $maxI$: Number of maximum iterations;
    $Nodes$: Number of nodes for the shallow layers of the residual network.
Output:
    Downscaled target variable with the minimum RMSE at fine resolution (1 km):
        $\hat{y}_m^{\mathrm{opt}}, m = 1, ..., r$ : Optimal estimate of target variable (downscaled M2GMI AOD) at 1km.
1. Initialization phase
    1.1 Direct derivation of the finely resolved samples ( $\hat{y}_m^{(0)}$ ) from coarse-resolution M2GMI AOD by overlaying;
    1.2 Construction of the network according to nodes and initialization of the network;
    1.3 $\mathbf{X}$ (predictors) and $\hat{\mathbf{y}}^{(0)}$ split into the training, validation and test samples.
2. Iteration
    iterator $t$=1 to $maxI$
        2.1 Adjust the estimators at the fine resolution in order to make their average equal to the value of the coarse-resolution cell that spatially covers them (i.e., equal to $Y_m$) based on Eq. (6);
        2.2 Train new network using the samples from $\mathbf{X}$ (as the predictors) and $\hat{\mathbf{y}}^{(t)}$, the adjusted estimators at fine resolution to obtain the output of the $t$+1 iteration, and calculate test $R^2$ and RMSE;
        2.3 Retrieve the minimum RMSE based on the $t$ test results;
        2.4 If the mean difference between $t$-1 and $t$ iterations smaller than the stopping criterion (<$SC$), break the loop; otherwise continue the iteration.
3. Return the minimum RMSE and its corresponding downscaled image as $\hat{y}_m^{\mathrm{opt}}, m = 1, 2, ...r$ .
```

Fig. 5.

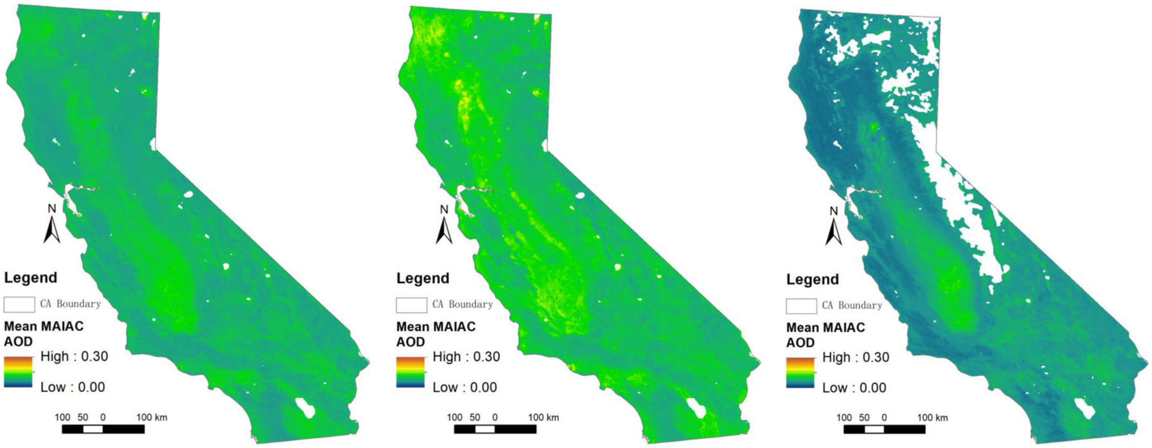
Downscaling algorithm for M2GMI AOD.



a. Average AOD coverage b. Summer AOD coverage c. Winter AOD coverage

Fig. 6.

Proportion (%) of available daily MAIAC AOD over California from 2000 to 2016: (a) overall; (b) summer (June-August); (c) winter (December-February).



a. Average AOD

b. Summer (Jun. to Aug.) AOD

c. Winter (Dec. to Feb.) AOD

Fig. 7.

Averages of multi-year (a), summer (b) and winter (c) daily MAIAC AOD across California from 2000 to 2016.

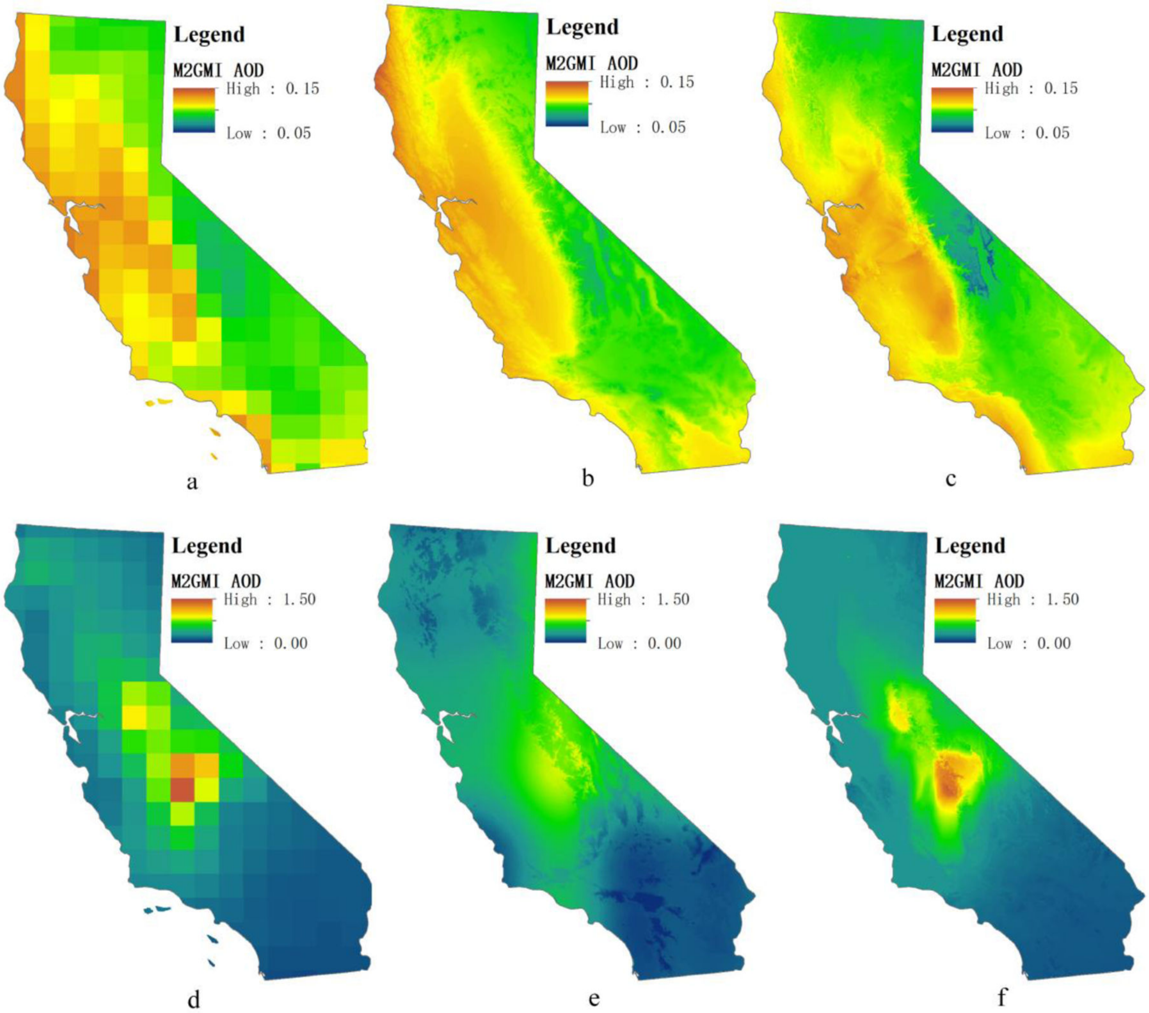


Fig. 8.

Comparison of the original M2GMI AOD (a and d), downscaled M2GMI AOD by GAM (b and e), and downscaled M2GMI AOD by deep residual network (c and f) for two weeks (05/01/2000–05/07/2000 for a, b and c; 09/14/2015–09/20/2015 for d, e and f).

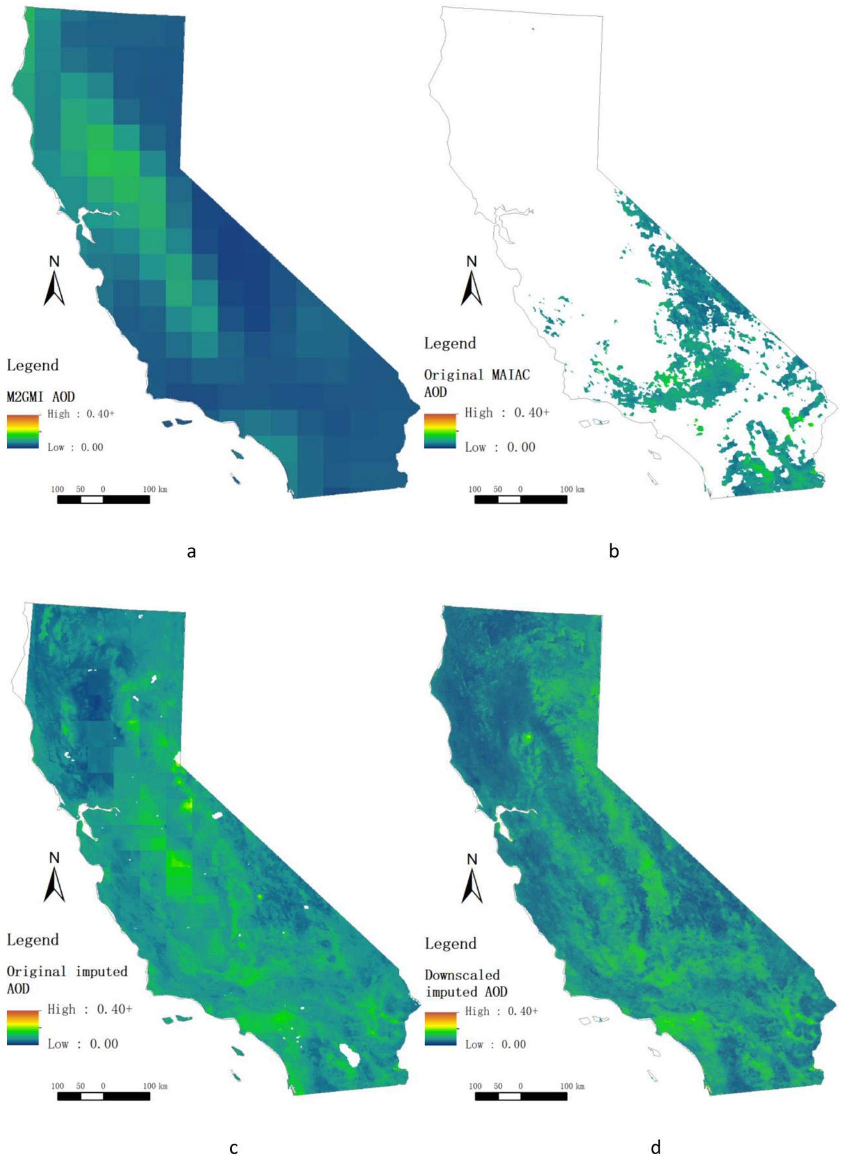
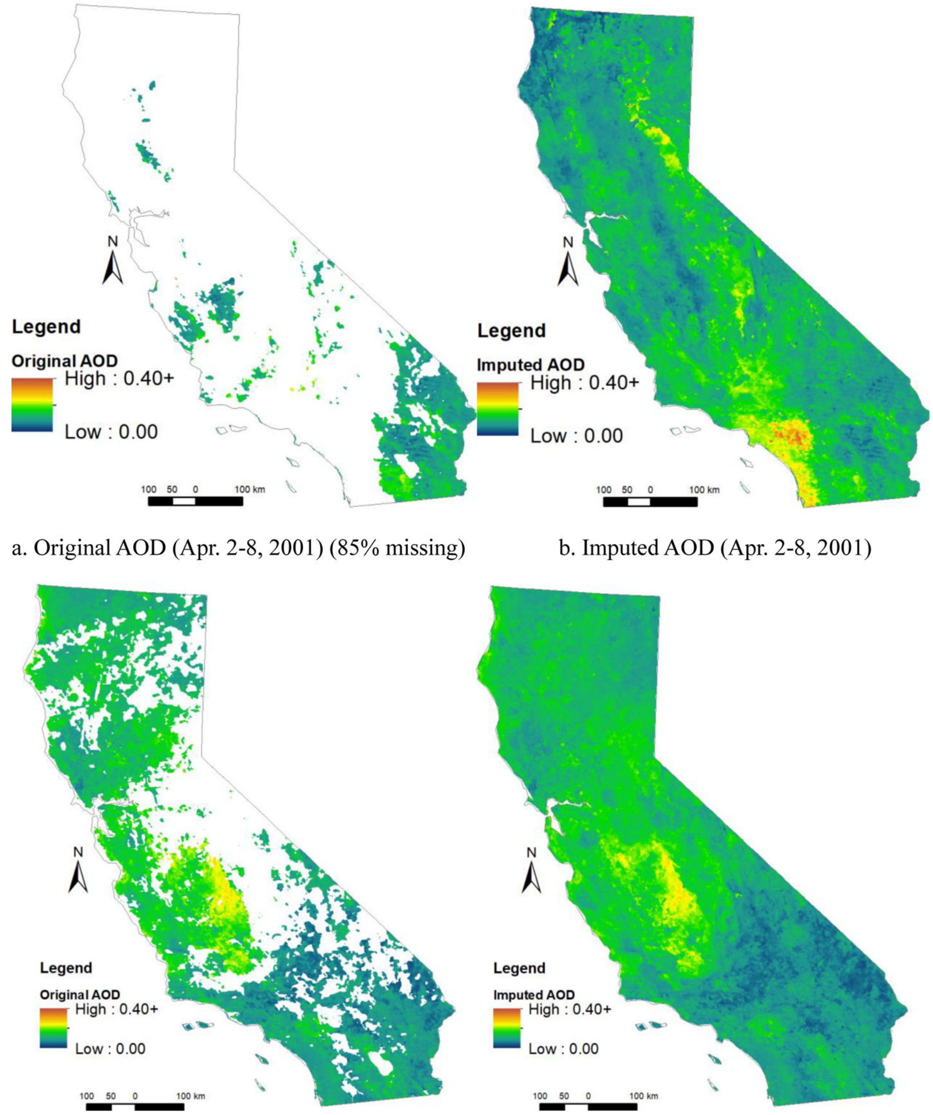


Fig. 9.

For a typical week (12/05/2016–12/11/2016), maps of grid surfaces of a) M2GMI AOD, b) MAIAC AOD with a significant proportion of missing observations, c) imputed MAIAC AOD with original 50 km M2GMI AOD, and d) imputed MAIAC AOD with downscaled 1 km M2GMI AOD

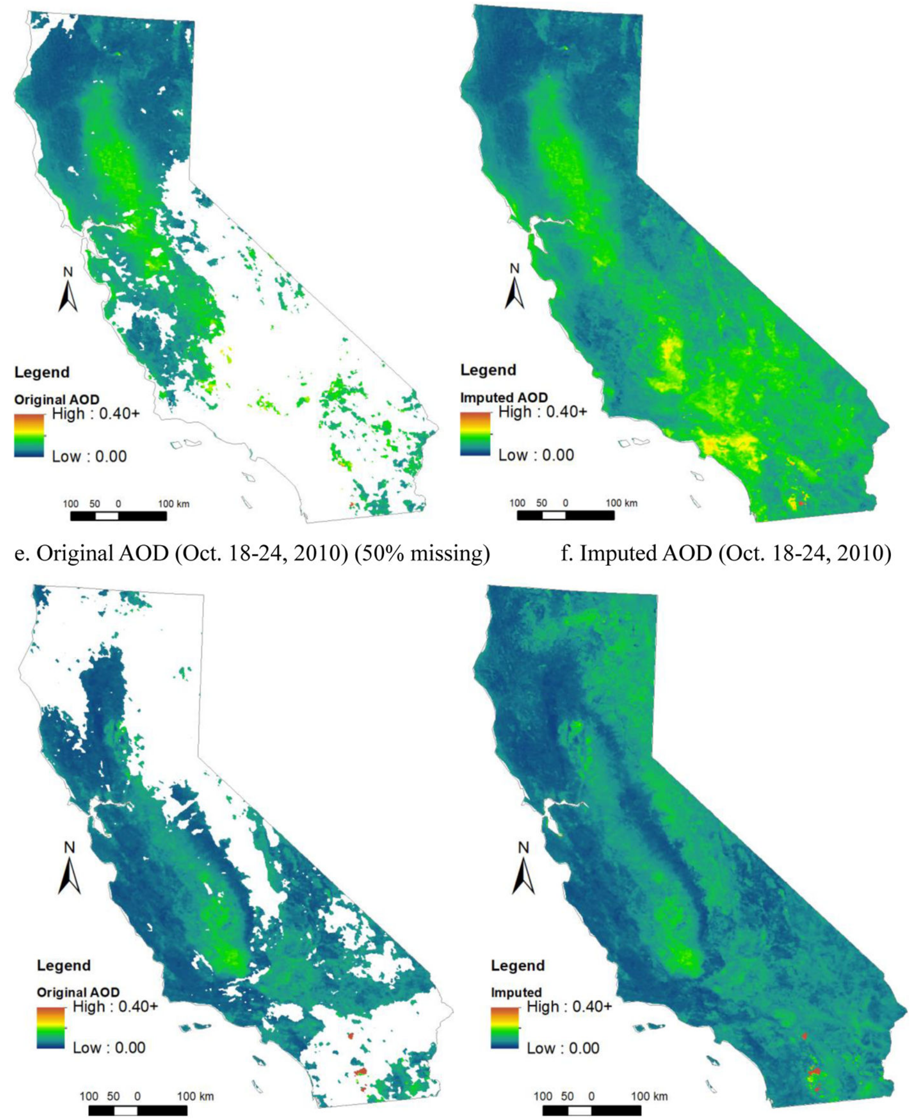


a. Original AOD (Apr. 2-8, 2001) (85% missing)

b. Imputed AOD (Apr. 2-8, 2001)

c. Original AOD (Jul. 28-Aug.3, 2003) (38% missing)

d. Imputed AOD (Jul. 28-Aug.3, 2003)

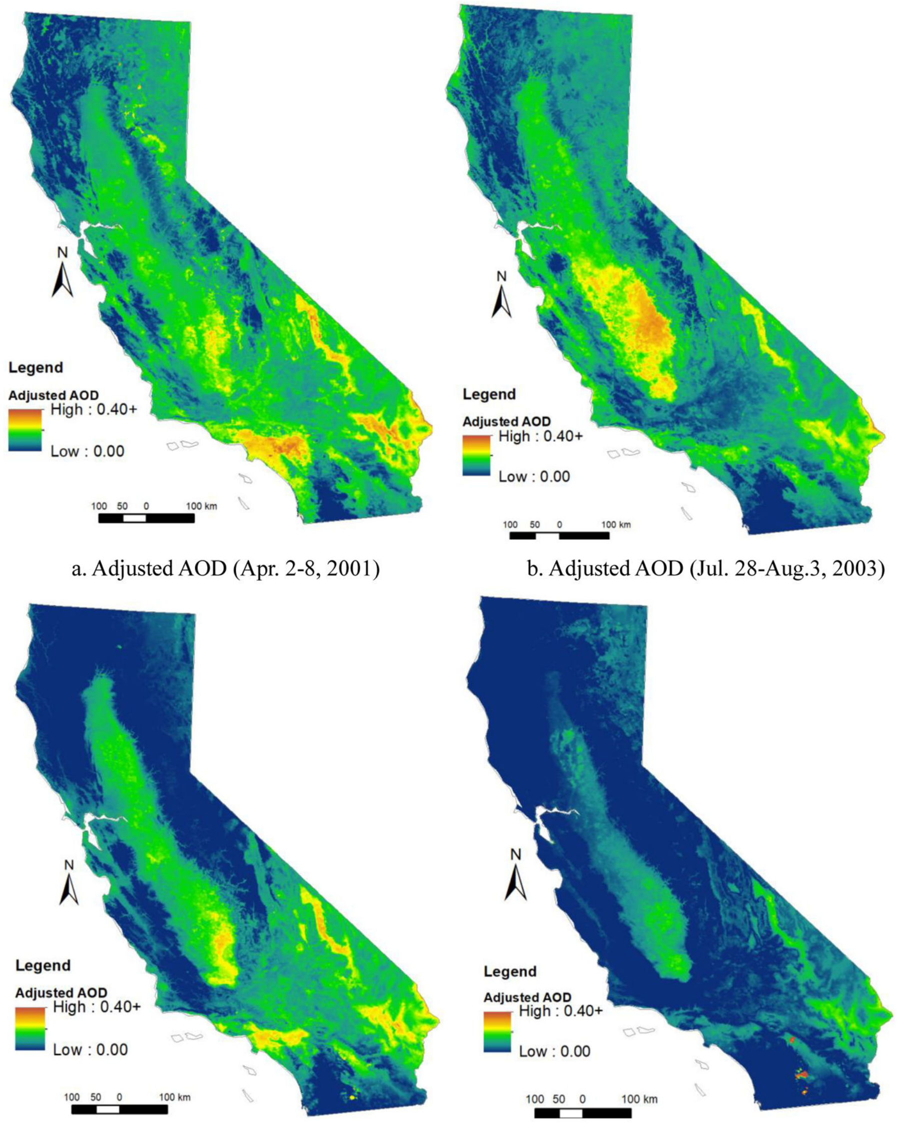


e. Original AOD (Oct. 18-24, 2010) (50% missing) f. Imputed AOD (Oct. 18-24, 2010)

g. Original AOD (Dec. 19-25, 2016) (45% missing) h. Imputed AOD (Dec. 19-25, 2016)

Fig. 10.

Grid surfaces of the missing (a, c, e and g) and imputed (b, d, f, and h) MAIAC AOD for four seasons of different years (a and b: spring of 2001; c and d: summer of 2003; e and f: autumn of 2010; g and h: winter of 2016).



a. Adjusted AOD (Apr. 2-8, 2001)

b. Adjusted AOD (Jul. 28-Aug.3, 2003)

c. Adjusted AOD (Oct. 18-24, 2010)

d. Adjusted AOD (Dec. 19-25, 2016)

Fig. 11.

Surfaces of bias-adjusted MAIAC AOD for four seasons of different years (a: spring of 2001; b: summer of 2003; c: autumn of 2010; d: winter of 2016)

Table 1.

Descriptive statistics of MAIAC AOD and input variables for California, 2000–2016

Group	Variable	Overall mean		Summer		Winter	
		mean	s.d.[a]	mean	s.d.	mean	s.d.
N	Sample number	1.5 billion		0.5 billion		0.2 billion	
Target variable	MAIAC AOD	0.084	0.017	0.100	0.024	0.062	0.018
	Monthly MAIAC AOD	0.081	0.039	0.100	0.056	0.063	0.025
Meteorology	Minimum temperature (°C)	7.880	7.570	15.070	5.870	1.480	4.950
	Maximum temperature (°C)	22.060	9.590	31.660	6.240	12.870	5.680
	Wind speed (m/s)	3.470	1.140	3.360	0.960	3.460	1.320
	Specific humidity (grams of vapor per kilogram of air)	0.005	0.002	0.007	0.002	0.004	0.001
	Shortwave radiation (w/m^2)	226.300	89.220	322.400	28.000	124.100	35.300
	Precipitation (mm m^{-2} h^{-1})	1.500	4.010	0.180	0.660	3.200	6.000
M2GMI	Daily mean AOD	0.093	0.073	0.120	0.110	0.062	0.038
Coordinates	Latitude	37.300	2.880	37.300	2.880	37.300	2.880
	Longitude	−119.170	3.130	−119.170	3.130	−119.170	3.130
Elevation	Elevation (m)	858.300	735.900	858.300	735.900	858.300	735.900

Note:

[a] s.d. Standard deviation.

Table 2.

Comparison of downscaling MERRA-2 GMI Replay Simulation AOD by GAM and deep residual network for two typical weeks and the average of the whole study period

Date	Model	Regular R^2	Regular RMSE	Validation[a] R^2	Validation RMSE	Independent test[b] R^2	Independent test RMSE
Averages for all weeks	Residual Network	0.89 (0.31–0.99)	0.0001 (6.95e-6–2.1e-2)	0.89 (0.23–0.99)	0.0001 (4.64e-6–0.16e-2)	0.89 (0.24–0.99)	0.008 (0.002–0.12)
	GAM	0.78 (0.22–0.98)	0.014 (0.003–0.22)	0.78 (0.22–0.98)	0.014 (0.003–0.22)	0.78 (0.22–0.99)	0.014 (0.002–0.12)
09/14/2015 – 09/20/2015	Residual Network	0.94	0.034	0.95	0.003	0.94	0.05
	GAM	0.66	0.14	0.66	0.14	0.65	0.14
05/01/2000 – 05/07/2000	Residual Network	0.87	2.52e-5	0.89	2.07e-5	0.89	0.004
	GAM	0.81	0.006	0.82	0.006	0.81	0.006

Note:

[a] Validation: 20% of the samples used for validation and also used to adjust the parameters during training

[b] Independent test: 20% of the samples used for independent test (the samples not used in training).

Table 3.

Performance of deep residual network (full and non-full models), regular network (full model) and GAM (full model) for weekly imputation of MAIAC AOD

Date of week	Model	Regular R^2	Regular RMSE	Validation R^2	Validation RMSE	Test R^2	Test RMSE
Averages for all weeks	Residual network (full model)[a]	0.94 (0.85–0.99)	0.007 (0.003–0.022)	0.94 (0.85–0.99)	0.007 (0.003–0.023	0.94 (0.84–0.99)	0.007 (0.004–0.023)
	Residual network (non-full model)[b]	0.86 (0.68–0.99)	0.01 (0.007–0.030)	0.86 (0.69–0.99)	0.01 (0.007–0.030)	0.84 (0.48–0.99)	0.01 (0.007–0.031)
	Regular network (full model)	0.78 (0.54–0.90)	0.013 (0.007–0.095)	0.80 (0.56–0.99)	0.011 (0.007–0.031)	0.73 (0.08–0.99)	0.015 (0.007–0.095)
	GAM (full model)	0.81 (0.57–0.93)	0.011 (0.008–0.032)	-	-	0.81 (0.56–0.93)	0.013 (0.009–0.037)
05/01/2000 – 05/07/2000	Residual network (full model)	0.92	0.008	0.92	0.008	0.92	0.008
	Residual network (non-full model)	0.85	0.010	0.85	0.010	0.82	0.011
	Regular network (full model)	0.75	0.012	0.76	0.011	0.70	0.013
	GAM (full model)	0.71	0.015	-	-	0.71	0.015
09/14/2015 – 09/20/2015	Residual network (full model)	0.98	0.012	0.98	0.012	0.98	0.012
	Residual network (non-full model)	0.94	0.021	0.94	0.021	0.94	0.021
	Regular network (full model)	0.92	0.023	0.93	0.022	0.94	0.022
	GAM (full model)	0.80	0.043	-	-	0.80	0.043

Note:

[a] full model means all the covariates used in training of the models

[b] non-full model means no use of monthly AOD mean but all the other covariates used in training of the models.

Table 4

Four validations of MAIAC AOD vs. AERONET AOD

Validation	*n*[a]	Pearson's correlation (no buffer/x km)[b]	R^2 (no buffer)	RMSE (no buffer)
Observed MAIAC AOD	2799	0.67 (no buffer) ; 0.71 (1 km); 0.73 (3 km); 0.74 (5 km); 0.75 (7 km); 0.75 (9 km)	0.44	0.06
Independent test point MAIAC AOD	737	0.81 (no buffer)	0.61	0.05
Imputed MAIAC AOD	122	0.60 (no buffer); 0.61 (1 km); 0.62 (3 km); 0.65 (5 km); 0.66 (7 km); 0.65 (9 km)	0.32	0.07
Combined observed and imputed MAIAC AOD	2921	0.69 (no buffer); 0.70 (1 km); 0.71 (3 km); 0.73 (5 km);0.73 (7 km); 0.74 (9 km)	0.45	0.06

Note:

[a:] Number of weekly samples

[b:] "(no buffering)" indicates a metric (correlation or RMSE) based on matched coincident pixel – site samples, and "(x km)" indicates a metric (correlation or RMSE) based on a spatial average within a circular buffer with a certain radius (e.g., 1 km, 3 km, 5 km, 7 km or 9 km radii) around each site.

Table 5.

Monthly variation for validation of weekly imputed MAIAC AOD with AERONET measurements.

Month	*n*[a]	Correlation		R^2		RMSE	
		Coincident[b]	9 km[c]	Coincident	9 km	Coincident	9 km
1	254	0.35	0.46	0.09	0.19	0.07	0.07
2	245	0.54	0.60	0.27	0.32	0.06	0.06
3	222	0.41	0.47	0.11	0.19	0.05	0.05
4	233	0.59	0.63	0.27	0.34	0.05	0.04
5	237	0.66	0.65	0.40	0.40	0.04	0.04
6	247	0.82	0.82	0.63	0.58	0.05	0.06
7	230	0.87	0.92	0.71	0.81	0.06	0.05
8	244	0.76	0.80	0.57	0.64	0.04	0.04
9	238	0.79	0.83	0.62	0.66	0.04	0.04
10	247	0.72	0.75	0.48	0.50	0.05	0.05
11	259	0.73	0.79	0.46	0.51	0.06	0.05
12	265	0.50	0.64	0.23	0.36	0.07	0.06

Note:

[a.] sample size

[b.] coincident pixel used

[c.] 9 km: average MAIAC AOD within 9 km buffer of AERONET site.