

© 2018 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

# Estimation of Power Plant Emissions with Unscented Kalman Filter

S. Metia, Q. P. Ha, H. N. Duc and M. Azzi

**Abstract**—Emissions from power plants constitute a major part of air pollution and should be adequately estimated. In this study, we consider the problem of estimating  $\text{NO}_x$  emissions of power plants by developing an inverse method to integrate satellite observations of atmospheric pollutant column concentrations with specie concentrations and direct sensitivities predicted by a regional air quality model in order to discern biases in the emissions of the pollutant precursors. Using this method, the emission fields are analyzed using a “bottom-up” approach with an inversion performed by an Unscented Kalman Filter (UKF) to improve estimation profiles from emissions inventories data for the Sydney Metropolitan area. The idea is to integrate information from the original inventory with tropospheric  $\text{NO}_2$  emissions estimated during one month from The Air Pollution Model-Chemical Transport Model (TAPM-CTM), and then, for validation, to compare the resulting model with satellite retrievals from the Ozone Monitoring Instrument (OMI) above the region. The UKF-based estimation of  $\text{NO}_2$  emissions shows better agreement with OMI observations, implying a significant improvement in accuracy as compared with the original inventories. Therefore, the proposed model is a promising tool for estimation of air emissions in urban areas.

**Index Terms**—Satellite data, UKF, Emissions inventories, Power plant, Nitrogen dioxide.

## I. INTRODUCTION

**N**ITROGEN OXIDES ( $\text{NO}_x = \text{NO} + \text{NO}_2$ ) are key precursors to tropospheric ambient ozone ( $\text{O}_3$ ), which are closely associated with adverse health effects, including respiratory diseases such as wheezing, coughing, colds, flu and bronchitis [1], [2]. It is therefore important to quantify and monitor  $\text{NO}_x$  emissions for air quality control in large cities. In this regard, point sources of this air pollutant can be identified by using satellite data retrieved from the Ozone Monitoring Instrument (OMI) [3], [4]. Emissions from power plants and large diffuse area sources have been estimated by using satellite record of atmospheric parameters, see, e.g., Krueger *et al.* [5], to extract sulfur dioxide emitted by volcanoes, combustion of fossil fuels or smelting of ores. Top-down nitrogen dioxide ( $\text{NO}_2$ ) column profiles can also be extracted from OMI data, see e.g., Bucsele *et al.* [6] and Mijling & van der A [7]. Validation on these estimations would however require a thorough comparison with other methods. For this,

satellite retrievals together with air quality models have been used to assess global and regional emissions, [8], [9].

Inverse modelling of ground level ozone concentrations has been the topic of many investigators such as [10], [11], but not much effort has been devoted to the development of a suitable model for prediction of spatio-temporal distributions of nitrogen oxides  $\text{NO}_x$  from emissions inventories. A similar work is reported in [12] to predict spatio-temporal distribution of columns of atmospheric carbon dioxide ( $\text{CO}_2$ ) using a gap-filling technique. There is a wide range of interest (both geochemical and geopolitical) in geographically explicit inventories of the sources and sinks of the greenhouse gas  $\text{CO}_2$ . It is a challenge to estimate sources and sinks in a spatially explicit context, and to best characterize the location and magnitude of emissions and sinks on regional scales [12], [13]. Most emissions inventories are developed by using the bottom-up approach, which is based on a combination of industrial, energy-consuming activities and source-specific or region-specific emissions factors [14]. Due to large uncertainties associated with the statistics, temporal profiles, grid allocations, natural sources, and human factors [15], [16], these emissions inventories require further improvements and refinements by using additional information such as station data, satellite data or other computational models, e.g. The Air Pollution Model with Chemical Transport Model (TAPM-CTM) [17].

Generally, there are many techniques to improve emissions inventories and predict severe air pollution in a region. In [18], an algorithm was developed to retrieve the haze aerosol optical thickness (HAOT) by using Moderate Resolution Imaging Spectrometer (MODIS) data to supplement the estimation with a global 3-D atmospheric chemical transport model (GEOS-Chem). In [19], a Kalman filter (KF) fusion algorithm has been applied to satellite data retrieved from the Total Ozone Mapping Spectrometer (TOMS)-Ozone Monitoring Instrument (OMI) combined with ground observations to estimate the error correlation region for UV-B monitoring. The Kalman filter inverse modelling technique was implemented on a Community Multiscale Air Quality (CMAQ)-based model to estimate  $\text{NH}_3$  emissions for the eastern United States [20]. Recently, an extended fractional Kalman filter has been proposed to the inverse problem for air pollutant emissions estimation and prediction for the Sydney basin in Australia [21].

The Global Ozone Monitoring Experiment-2 (GOME-2) is one of the new-generation European instruments carried on MetOp (a series of three polar orbiting meteorological satellites) and will continue the long-term monitoring of atmospheric ozone started by GOME on ERS-2 (European Remote Sensing) and SCIAMACHY (SCanning Imaging Absorption

S. Metia and Q. P. Ha are with Faculty of Engineering & IT, University of Technology, University of Technology Sydney, Broadway NSW 2007, Sydney, Australia (e-mail: Metia.Santanu@student.uts.edu.au and Quang.Ha@uts.edu.au).

H. N. Duc is with Office of Environment and Heritage, Lidcombe NSW 2141, Sydney, Australia (e-mail: Hiep.Duc@environment.nsw.gov.au).

M. Azzi is with CSIRO Energy, North Ryde NSW 2113, Sydney, Australia (e-mail: merched.azzi@csiro.au).

Manuscript received March 3, 2018.

SpectroMeter for Atmospheric CHartography) on Envisat. The more advanced GOME-2 is set to make a significant contribution towards climate and atmospheric research, whilst providing near real-time data for use in air quality forecasting. Different gases in the atmosphere absorb different wavelengths of light. The GOME-2 scanning spectrometer is designed to exploit this fact by capturing light reflected from the Earth's surface and atmosphere and splitting it into different wavelengths to reveal absorption lines, which correspond to certain gases present in the observed sample. GOME-2 will map concentrations of atmospheric ozone as well as nitrogen dioxide, sulphur dioxide, other trace gases and ultraviolet radiation. These profiles are representative of the lowermost 50 km of the Earth's atmosphere. These data are crucial for monitoring stratospheric ozone and atmospheric pollutants to keep a check on the health of the Earth's atmosphere.

In this paper, the problem of estimating power plant emissions is considered. The ultimate goal is to understand the environmental impact of isolated power plants towards achieving sustainable energy usage [22]. Since emissions data, considered as time series, are highly nonlinear, linearized estimators such as the extended Kalman filter do not yield reliable performance. Here, by formulating the emissions prediction as the state estimation problem for a multi-dimensional continuous-time Gaussian process, an Unscented Kalman filter (UKF) for the air-pollutant emissions model can be used to deal with nonlinearities [23]. The filter is also coupled with a Matérn class of covariance functions to take into account spatio-temporal distributions and smoothing of the correlation [24]. As such, the proposed approach can deal with cases of missing or non-uniformly sampled data, towards solving the inversion problem of air pollutant emissions.

The remainder of this paper is organised as follows. Following the introduction, Section II presents the emissions data management system and UKF estimation scheme. Section III provides the estimation results for NO<sub>2</sub> emissions using the proposed UKF method with TAPM-CTM. Section IV presents the estimated emissions and the satellite-based retrievals from OMI data over the power plant region of the Greater Metropolitan Region (GMR) in the New South Wales (NSW) state. The proposed model is validated through comparison to confirm its accuracy in prediction of air pollutant emissions, here coming predominantly from power plants. Finally, a conclusion is drawn in Section V.

## II. EMISSIONS INVENTORIES AND UNSCENTED KALMAN FILTER

Inverse pollution estimation involves the identification of the emissions sources and the amount of their emissions in environment. Air quality plans are nowadays being developed in urban and regional areas where air quality does not comply with the limit values established by authorities. Air quality models are essential tools to support policy formulation by evaluating the possible impact of local and regional emission abatement options on air quality and human health. Therefore, there is a need for better understanding the air quality model uncertainties and ensure they are fit-for-purposes. The

uncertainties mostly relay on the input data, such as meteorology, boundary conditions and, emissions. This identification process is typically carried out by using the bottom-up approach [25], resulting in databases called emissions inventories. For estimation of air pollutant distributions in a region, the reliability of emissions inventories is crucially important in air quality management. This section presents the UKF and Matérn covariance function based approach to deal with nonlinearities and uncertainties associated with emissions inventories.

### A. Emissions data management system

In this paper, the Emissions Data Management System (EDMS v2.0) [26] is used to generate data of air pollutant emissions, i.e. the original or *priori* emissions inventories, for GMR NSW, Australia. To determine the spatio-temporal profiles of emissions in those inventories, map coordinates are assigned for industrial and commercial controlled point sources. These coordinates are based on 1-km by 1-km grid cells for natural (biogenic), domestic, off-road and on-road area sources as well as industrial and commercial uncontrolled fugitive sources. Emissions are then generated for each month, day of week, and hour of day, taking into account a variety of emitting activities that occur in transport and industries. For example, the transport sector comprises emissions from registered vehicles operating on public roads, such as cars, motorcycles, trucks and buses, both petrol- and diesel-fuelled. These emissions are allocated across the road network according to data of vehicle's kilometers travelled. Since uncertainties associated with emission inventories are hardly avoidable, a good model is required to adequately relate emissions inventories with air pollution emissions.

### B. UKF scheme

Consider the air-pollutant dispersion model and Matérn covariance function as described in [21]:

$$\begin{aligned} \frac{dx(t)}{dt} &= \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -\lambda^3 & -3\lambda^2 & -3\lambda \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} w(t) \\ y_i &= [1 \ 0 \ 0] x(t_i) + \varepsilon_i, \end{aligned} \quad (1)$$

where  $x(t)$  is the state vector,  $w(t)$  is white noise,  $\lambda$  is a coefficient depending on the correlation length and smoothness of the process, and  $\varepsilon_i$  is an independent zero-mean Gaussian random variable with variance  $\sigma_i^2$  for  $i = 1$  to  $N$  data points [27]. The dynamics of air-pollutant are considered as  $x(t)$  and measurement model output are  $y_i$ . To improve emission inventories, an UKF will be developed for state estimation in face of the system nonlinearity. Unlike the extended Kalman filter (EKF) which uses the first-order approximation of the nonlinear system, UKF represents a derivative-free alternative with lesser computational complexity [28]. Generally, an unscented transform (UT) can be used to estimate the distribution of a *posteriori* state, taking into account nonlinearities of the process [29], whereby UKF is expected to yield more accurate estimates compared to EKF owing to the capability

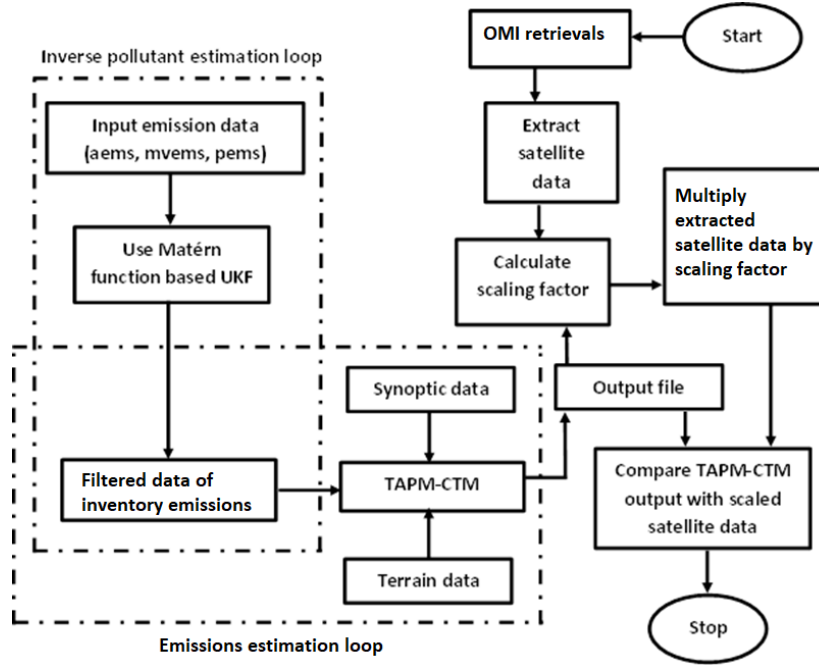


Fig. 1: Flowchart of the proposed emissions estimation using Matérn function based UKF with comparison of data extracted by satellite.

of propagating mean and covariance information through the nonlinear transformations [23].

In our system, a Gaussian approximation of the joint distribution of random variables  $x$  and  $y$  is adopted for the unscented transform:

$$\begin{aligned} x &\sim \mathcal{N}(m, P) \\ y &= g(x), \end{aligned} \quad (2)$$

where  $x \in \mathbb{R}^n$ ,  $y \in \mathbb{R}^m$  and  $g : \mathbb{R}^n \mapsto \mathbb{R}^m$  is a general nonlinear function. The UT main idea is to form a fixed number of deterministically-chosen sigma-points to appropriately capture the mean value of  $x$  as well as the covariance of the original random process. These sigma-points are then propagated through the nonlinear process so that the mean and covariance of the transformed variables can be estimated in the presence of nonlinearities. Specifically, the unscented Kalman filter (UKF) [30] is a discrete-time system represented by the following equations:

$$\begin{aligned} x_k &= f_d(x_{k-1}, k-1) + q_{k-1} \\ y_k &= h_d(x_k, k) + r_k, \end{aligned} \quad (3)$$

where  $x_k \in \mathbb{R}^n$  is the state,  $y_k \in \mathbb{R}^m$  is the measurement,  $q_{k-1} \in \mathbb{R}^n$  is a Gaussian process noise  $q_{k-1} \sim \mathcal{N}(0, Q_{k-1})$ , and  $r_k \in \mathbb{R}^m$  is a Gaussian measurement noise  $r_k \sim \mathcal{N}(0, R_k)$ . The mean and covariance of the initial state  $x_0$  are  $m_0$  and  $P_0$ , respectively. Using the UT matrix form described in [28], the prediction and update steps of UKF to enhance the emission inventory are conducted in the following steps:

- *Prediction* :

Compute the predicted state mean  $m_k^-$  and the predicted covariance  $P_k^-$  as:

$$\begin{aligned} [m_k^-, \tilde{P}_k^-] &= UT(f_d, m_{k-1}, P_{k-1}) \\ P_k^- &= \tilde{P}_k^- + Q_{k-1}. \end{aligned} \quad (4)$$

- *Updating* :

Compute the predicted mean  $\mu_k$  and covariance of the measurement  $S_k$ , and the cross-covariance of the state and measurement  $C_k$ :

$$\begin{aligned} [\mu_k, \tilde{S}_k, C_k] &= UT(h_d, m_k^-, P_k^-) \\ S_k &= \tilde{S}_k + R_k. \end{aligned} \quad (5)$$

The filter gain  $K_k$ , state mean  $m_k$  and covariance  $P_k$ , conditional to the measurement  $y_k$ , are then computed as:

$$\begin{aligned} K_k &= C_k S_k^{-1} \\ m_k &= m_k^- + K_k [y_k - \mu_k] \\ P_k &= P_k^- - K_k S_k K_k^T. \end{aligned} \quad (6)$$

The unscented transform is used for forming the Gaussian approximation using  $2n + 1$  sigma points from the columns of the  $n \times n$  matrix [30]. The parameter  $\lambda$  is a scaling parameter defined as

$$\beta = \alpha^2(n + \kappa) - n. \quad (7)$$

The positive constants  $\alpha$  and  $\kappa$  are used as parameters of the method.

### III. ENHANCED EMISSIONS ESTIMATION USING UKF

This section describes the proposed model for estimation of air pollutant emissions and compares the results of the

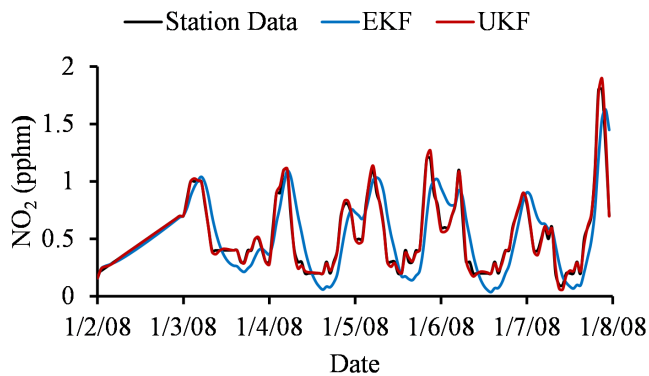


Fig. 2: NO<sub>2</sub> level at Liverpool station (latitude -33°55'58", longitude 150°54'21" and altitude 22) from 2<sup>nd</sup>-7<sup>th</sup> January, 2008.

estimation from using UKF and EKF. Here, after using a suitable Matérn covariance function for smoothing the correlation of the spatio-temporal emissions distributions, the inventories are incorporated into chemical transport models to take into account information of the terrain and data of the synoptic distributions of meteorological conditions over the region.

#### A. Air quality model

The proposed model for emission estimation using the Matérn function-based UKF is shown in the flowchart of Fig. 1 with two loops: the first loop corresponds to the inverse pollutant estimation and the second one is for emissions estimation. Station data are used to determine the order for the Matérn covariance function. Here, a 3<sup>rd</sup> order of the covariance function is obtained by using a genetic algorithm [21]. After filtering emissions data with the Matérn function-based UKF, the air pollutant profiles are obtained, considering the chemical transport, synoptic data and terrain data. For this, we use the TAPM-CTM [17] as the grid model, applied for the whole Australian continent with nested 3/1 km resolution grids, from the earth's surface upward to a height of approximately 4 km (vertical grid structure). The model solves the momentum equations for wind velocity components in the horizontal direction. Its meteorological component predicts the flow of sea breezes and terrain-induced circulations, given the larger scale synoptic fields data. In this work, inventory data of air pollutants emitting from areas, points and motor vehicles are generated by EMDS v2.0 and organised into area emissions (*aems*), point emissions (*pems*), and vehicle emissions (*mvems*) [21]. These files contain 16 species including NO and NO<sub>2</sub>. These data are filtered by using the proposed Matérn function-based UKF, then, together with synoptic and terrain data, they form the input to the TAPM-CTM to obtain air pollutant profiles. The output is then compared with data extracted from satellite retrievals for model validation after using scaling factors.

#### B. Station measurements

Measurements collected at monitoring stations serve as the ground's benchmarking data for evaluation of the TAPM-CTM output data to measure the accuracy of output data. Observation data for O<sub>3</sub>, NO<sub>2</sub>, CO, SO<sub>2</sub>, PM<sub>10</sub> and PM<sub>2.5</sub> obtained from 20 sites in the GMR are integrated to form the basis for comparison of spatio-temporal profiles of these air pollutants. Hourly pollutant concentration measurements are collected at monitoring stations and made available online by The Office of Environment and Heritage in New South Wales (NSW) [31], providing a basis for air quality control for the Sydney metropolis. In Fig. 2, a recorded profile of NO<sub>2</sub> in real-time is shown together estimated profiles using UKF and EKF.

#### C. Comparison between UKF and EKF on emissions estimation

Results from the TAPM-CTM air quality model are compared with air pollutant profiles generated by using the original and filtered inventories. With uncertainties attributed to air pollution emissions, the estimation performance is affected if the dynamic system is subject to high nonlinearities, unknown external sources or parameter variations. López-Aparicio *et al.* [32] show discrepancies in nitrogen oxides (NO<sub>x</sub>) and particulate matter (PM<sub>2.5</sub> and PM<sub>10</sub>) when evaluating both total and sectorial emissions. These discrepancies are associated with the assumptions made for the allocation of emissions. Emission inventories are developed at local, regional and national scales, with methods that very much depend on the purpose, emission source intensity and input data availability. To overcome this problem, the UKF, being derivative-free and capable of propagating the mean and covariance through an unscented transform, is considered as a better option than EKF. For comparison, both filters were used to estimate the temporal distribution of the same air pollutant at a station site. The obtained results indicate that the estimation follows the observations, as shown in Fig. 2. Therein, NO<sub>2</sub> is evident near emission sources such as coal-fire power plant and motor vehicle exhaust. However, in terms of estimation accuracy. The mean squared error (MSE) of the NO<sub>2</sub> profile using EKF and UKF are respectively 0.0427 and 0.0012, with the latter being more accurate, as expected.

The spatial concentration plots of NO<sub>2</sub> distributions over the GMR in New South Wales, as estimated by using the original and the improved inventories with UKF, are shown respectively in Figs. 3a and 3b in summer time (January 2008) as well as in Figs. 4a and 4b in winter time (July 2008). By referring to the location and ratings of power plants across the state, provided in Table I, the impact of power plant emissions can be judged by looking at the highest concentration of NO<sub>2</sub> in the figures. Fig. 5 shows map of coal based power stations and monitoring stations across the NSW. In addition, as it can be seen therein, the overestimation and underestimation of the spatial distributions of NO<sub>2</sub> emissions, associated with the original inventory in regions affected by power plants, have been rectified by using the Matérn function based UKF. To be more specific, Fig. 6a and Fig. 6b show

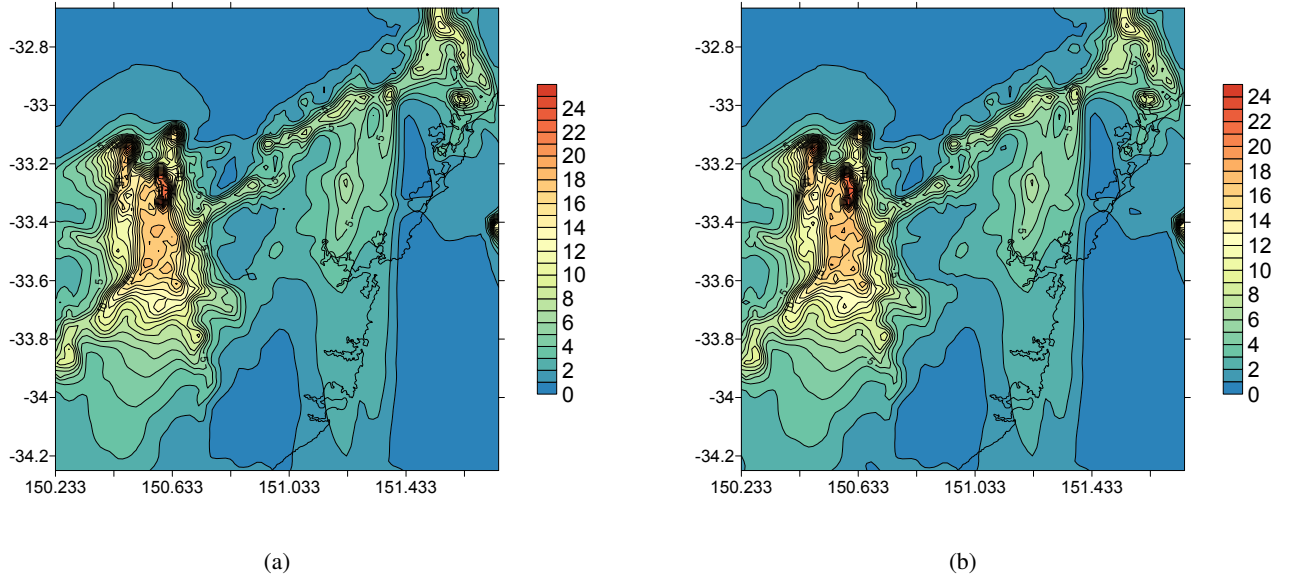


Fig. 3: Estimated  $\text{NO}_2$  concentrations (ppb) over NSW for January 2008 using TAPM-CTM: (a) original inventory, and (b) with UKF.

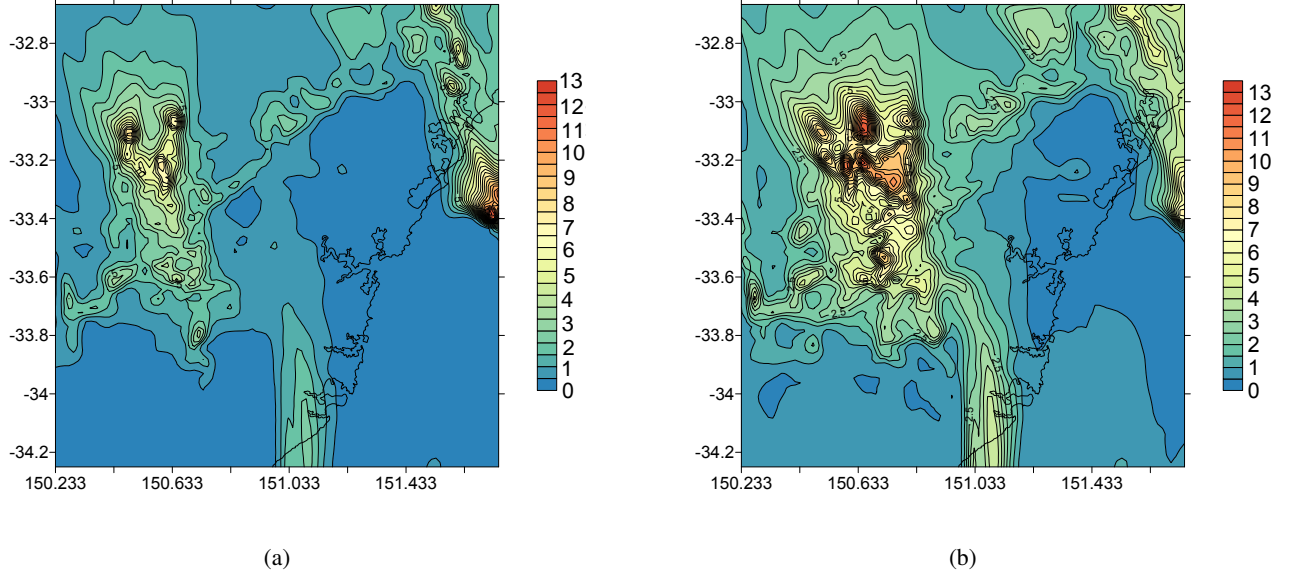


Fig. 4: Estimated  $\text{NO}_2$  concentrations (ppb) over NSW for July 2008 using TAPM-CTM: (a) original inventory, and (b) with UKF.

the difference in  $\text{NO}_2$  concentrations obtained by using the original inventory and the UKF-based estimation for January and July months respectively, where it can be observed of the  $\text{NO}_2$  emissions concentrations are slightly overestimated over the ocean in the eastern part of the GMR while they are inherently underestimated over the interested power plant regions, even more seriously in the winter (July in the southern hemisphere).

#### IV. COMPARISON BETWEEN TAPM-CTM-ESTIMATED AND OMI-RETRIEVED $\text{NO}_2$ COLUMNS

In this section, we compare OMI tropospheric  $\text{NO}_2$  columns with ground based TAPM-CTM output of  $\text{NO}_2$  pollutant profile. The OMI instrument on board the NASA/EOS-Aura satellite, a nadir-viewing imaging spectrometer, provides in-

TABLE I: MAJOR POWER STATIONS IN NSW (COAL-BASED)

Power station	Location	Nearest monitoring station	Installed capacity (MW)	Latitude (Degree)	Longitude (Degree)
Bayswater	Hunter	Beresfield	2,720	-32.8765°	151.2290°
Eraring	Lower Hunter	Wallsend	2,640	-33.0664°	151.5180°
Liddell	Hunter	Muswellbrook	2,080	-32.4028°	151.0183°
Mount Piper	Central West	Bathurst	1,400	-36.3817°	148.4178°
Vales Point	Central Coast	Wyong	1,320	-33.1489°	151.4884°
Wallerawang	Central West	Bathurst	1,000	-33.4113°	150.0649°

formation on the properties of aerosols and clouds as well as global levels of atmospheric species such as ozone ( $\text{O}_3$ ), Nitrogen dioxide  $\text{NO}_2$ , Sulfur dioxide  $\text{SO}_2$ , Chlorine dioxide ( $\text{ClO}_2$ ), Bromine oxide ( $\text{BrO}$ ), and Formaldehyde ( $\text{HCHO}$ ) on



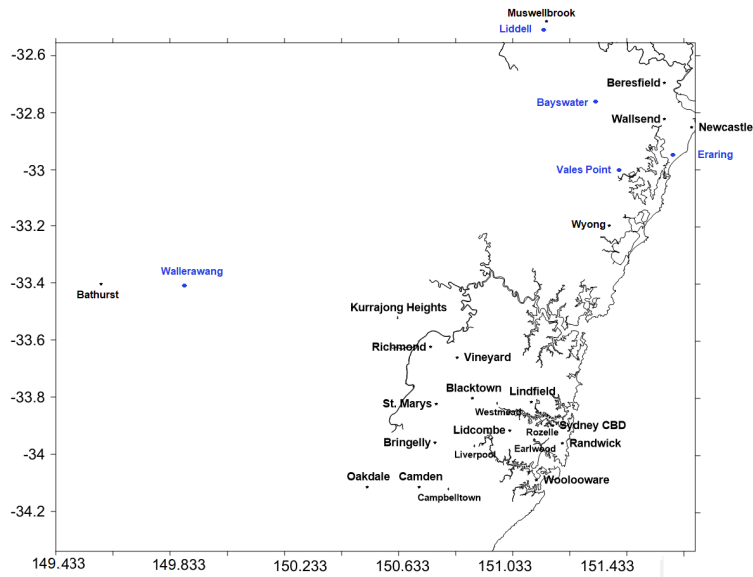


Fig. 5: Map of monitoring stations (black) and coal based power stations (blue) across the NSW.

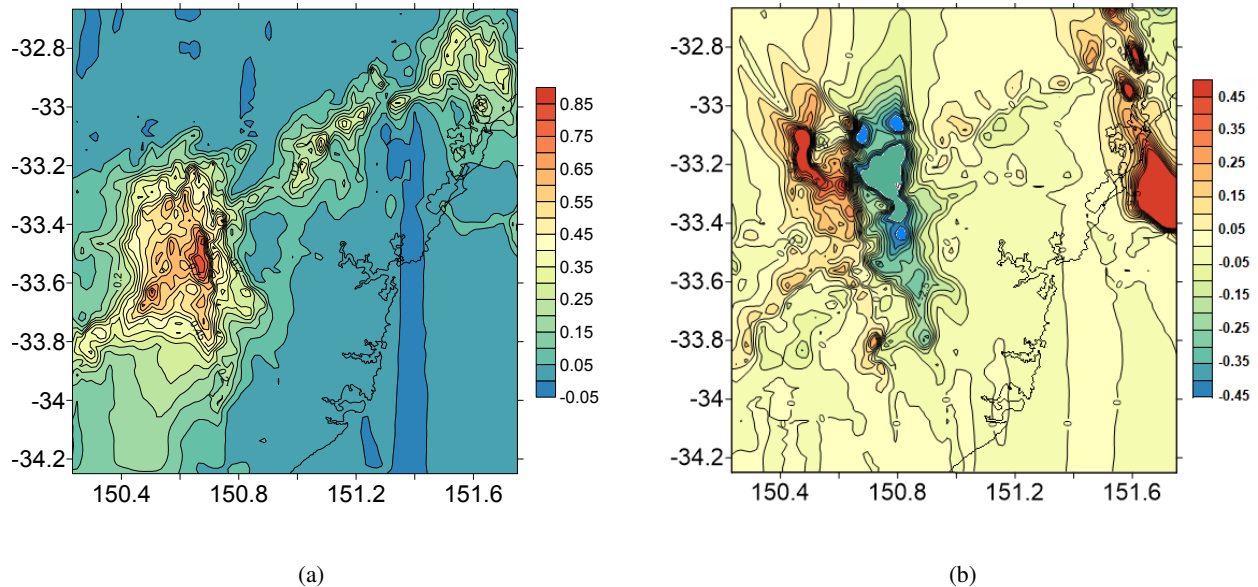


Fig. 6: Difference between the original inventory and UKF-based estimation of  $\text{NO}_2$  concentrations (ppb) for (a) January 2008, and (b) July 2008.

a daily basis.

#### A. OMI retrievals

The Ozone Monitoring Instrument (OMI) was launched on NASA's Aura satellite in July 2004 and has been providing measurements of ultraviolet and visible radiation with a spatial resolution of up to 13 by 24  $\text{km}^2$  [33]. In this work, data obtained from the OMI with a maximum resolution of 13 by 24  $\text{km}^2$  are used to create  $\text{NO}_2$  distribution maps over Australia at  $0.1^\circ$  resolution. These maps can be further compared with  $\text{NO}_2$  data from surface concentrations measured at station sites, as adopted in [34]. The results are used to validate the proposed UKF-based method for estimation of the total  $\text{NO}_2$  emissions in NSW considering particularly locations of the power plants.

#### B. Satellite data estimation

In this work, data are downloaded from Giovanni (<http://giovanni.gsfc.nasa.gov>), which provides an online tool for visualization, data extraction, exploration, and data analysis from NASA Earth Science. Spatial plots of  $\text{NO}_2$  total column in January 2008 for the NSW region are shown in Fig. 7a. It can be interpreted that fossil fuel burning and forest fires have led to high emissions of  $\text{NO}_2$  across the state during mid-summer. Moreover, in areas with high  $\text{NO}_2$  emission over the state, a good coincidence can be observed between satellite data and the UKF-based estimation method. Fig. 7b shows spatial plots of  $\text{NO}_2$  total column the NSW region in July 2008. Due to meteorological effects, the  $\text{NO}_2$  distribution tended to shift towards the sea. A comparison of the satellite

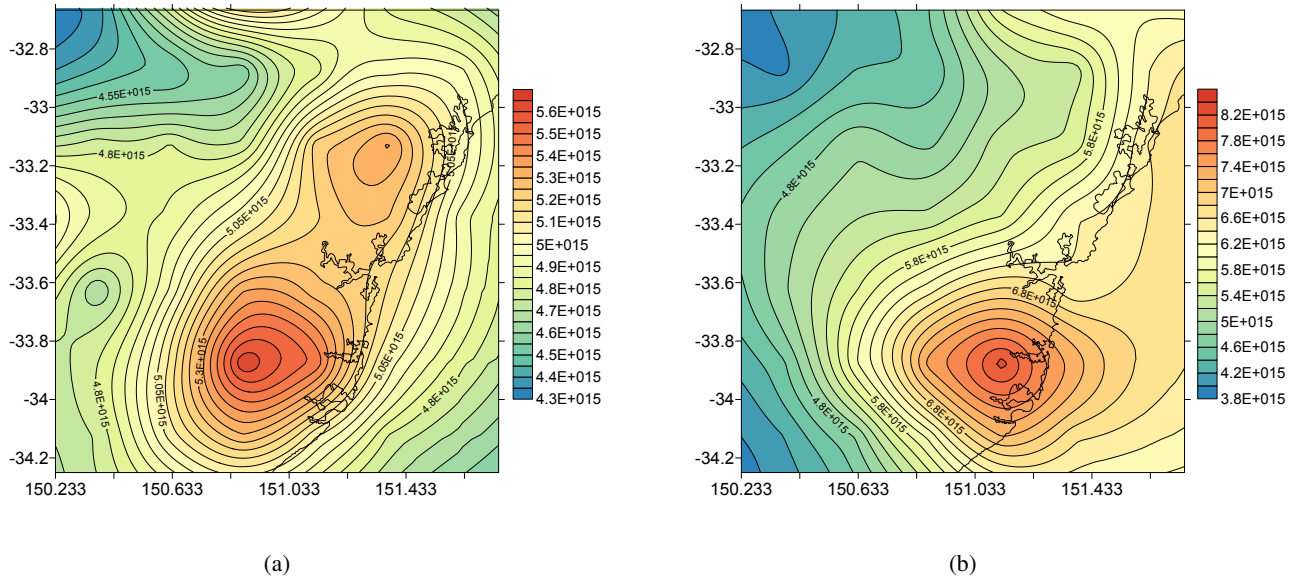


Fig. 7: Spatial plots of NO<sub>2</sub> emissions (molec/cm<sup>2</sup>) from satellite data over NSW for (a) January 2008, and (b) July 2008.

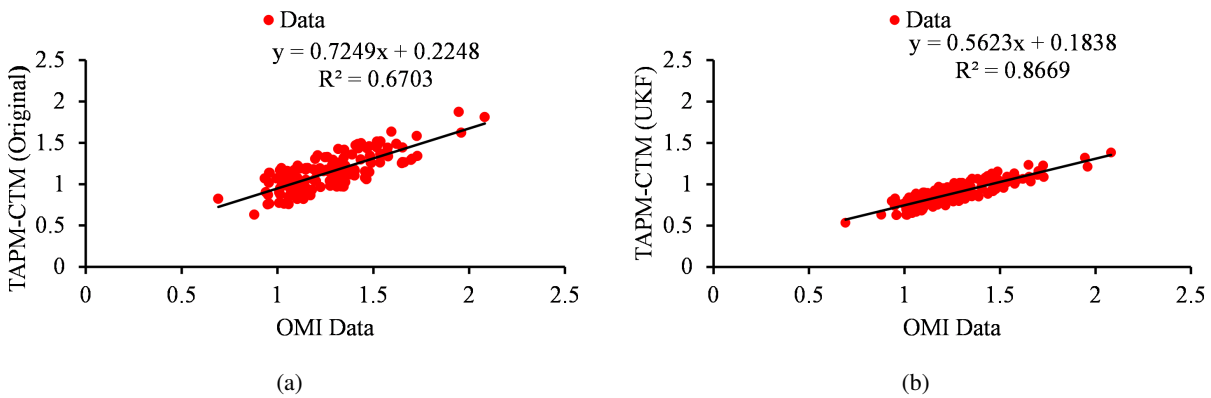


Fig. 8: Scatter plots of TAPM-CTM and the OMI-derived surface NO<sub>2</sub> for January 2008, TAPM-CTM retrievals are shown in red circles: (a) original inventory, and (b) with UKF.

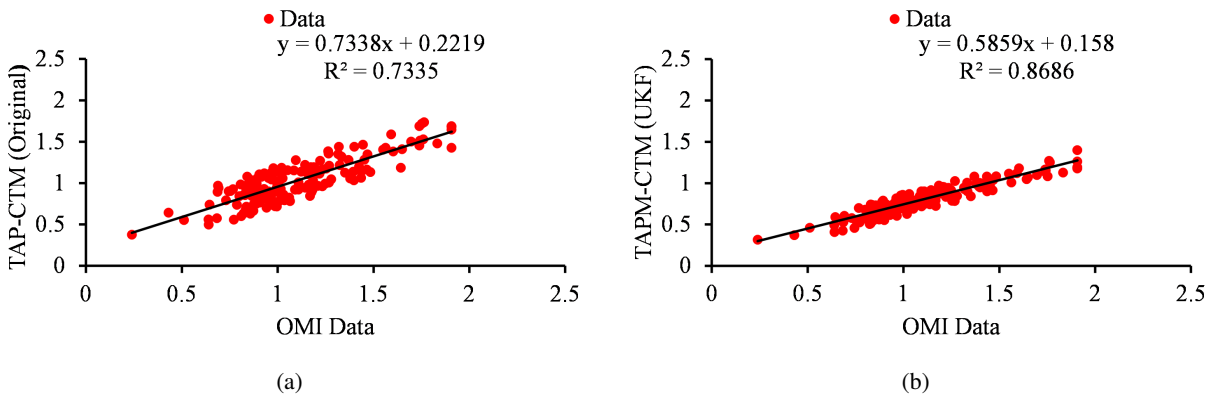


Fig. 9: Scatter plots of TAPM-CTM and OMI-derived surface of NO<sub>2</sub> for July 2008, TAPM-CTM retrievals are shown in red circles: (a) original inventory, and (b) with UKF.

retrievals with the estimation shown in Fig. 4b indicates that the UKF-based model has overcome the underestimation of NO<sub>2</sub> concentrations from the original inventory to emphasize the contribution of power plant sources to emissions, espe-

cially in the winter. However, there still exists a difference in NO<sub>2</sub> concentrations between estimation by the UKF and remotely sensing from satellite data. This can be explained by two reasons. First, the UKF-based estimation could not include



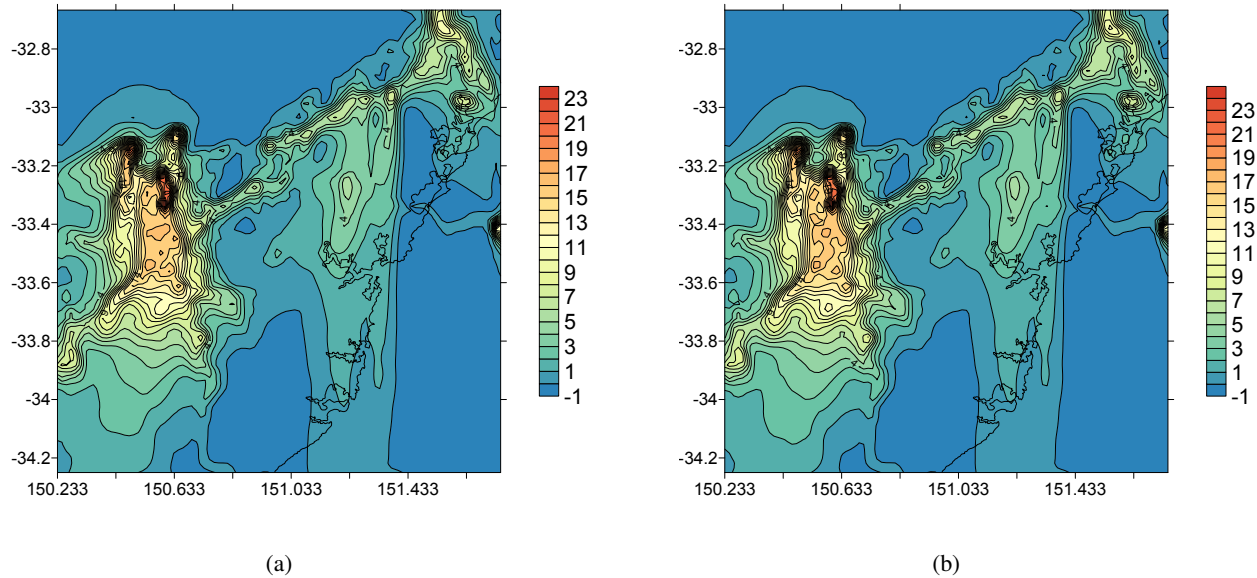


Fig. 10: Difference between TAPM-CTM and OMI of  $\text{NO}_2$  concentrations (ppb) for January 2008: (a) original inventory, and (b) with UKF.

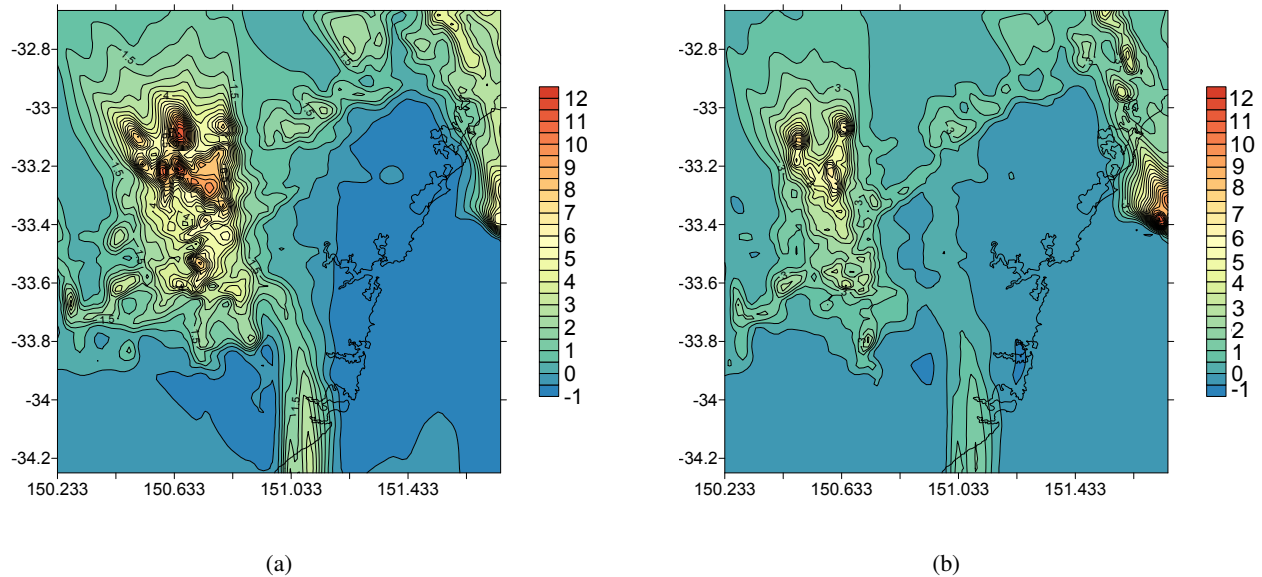


Fig. 11: Difference between TAPM-CTM and OMI of  $\text{NO}_2$  concentrations (ppb) for July 2008: (a) original inventory, and (b) with UKF.

all factors influencing  $\text{NO}_2$  distributions such as emissions aloft from lightning sources or airplanes. Second, the UKF-based method estimates pollutant concentrations at the surface level while satellite data are obtained in the troposphere. For better comparison, it is essential to use scaling factors considering emissions at the ground level.

### C. Surface concentration comparison

For model-model comparison at the ground level, the TAPM-CTM hourly forecasts for  $\text{NO}_2$  concentrations are evaluated against OMI  $\text{NO}_2$  observations. Five power plant sites (Wallerawang, Vales Point, Eraring, Bayswater and Lindell) and five monitoring station sites (Muswellbrook, Beresfield, Wallsend, Bathurst and Wyong) are considered for analysis. To this end, surface  $\text{NO}_2$  concentrations are inferred from the

satellite observations by calculating a scaling factor, defined as the ratio of the average *in situ* surface  $\text{NO}_2$  concentrations to the average OMI column abundance levels [35]. This ratio varies between seasons due to changes in natural  $\text{NO}_2$  emissions. Besides, there is a mismatch between TAPM-CTM grid numbers (57330 grid cells,  $1 \times 1 \text{ km}^2$  each) and OMI grid numbers (40 grid cells,  $13 \times 24 \text{ km}^2$  each). These satellite observations are not “real” or “true” values, having different vertical sensitivities at different altitudes in the atmosphere. To consider this vertical sensitivity of the satellite observations, the mass balance method should be introduced into comparison studies between TAPM-CTM-simulated and satellite retrieved tropospheric  $\text{NO}_2$  columns [36]. Han *et al.* [37] show scaling factors vary according to seasons. The scaling factors are calculated according to the mass balance method

TABLE II: STOCHASTIC GOODNESS-OF-FIT CRITERIA FOR THE ORIGINAL INVENTORY AND UKF ESTIMATION

Month	Type of inventory	MSE	R <sup>2</sup>	p-value
January	Original	0.1519	0.6703	$1.8 \times 10^{-4}$
	UKF	0.0384	0.8669	$1.2 \times 10^{-4}$
July	Original	0.1125	0.7335	$2.1 \times 10^{-4}$
	UKF	0.0320	0.8686	$1.7 \times 10^{-4}$

[36], [37]. As a result, the scaling factors are  $2.5700 \times 10^{16}$  molec/cm<sup>2</sup> for original inventory and  $2.5240 \times 10^{16}$  molec/cm<sup>2</sup> for UKF inventory during January 2008. Similarly, these values are  $1.9458 \times 10^{16}$  molec/cm<sup>2</sup> for original inventory and  $1.8371 \times 10^{16}$  molec/cm<sup>2</sup> for UKF inventory during July 2008.

The performance of emission estimation with the proposed UKF model is now evaluated by comparing with satellite data, as summarized in Table II. From statistical goodness-of-fit criteria such as the mean-squared error (MSE), the coefficient of determination R<sup>2</sup> between the measurement and prediction, as well as the *p*-value, both the original inventory and UKF correction indicate good estimation results with the UKF-based prediction outperforming the original inventory for NO<sub>2</sub> emissions. As depicted in Fig. 8 and Fig. 9 for the regression lines between satellite data retrieved from the Giovanni server and TAPM-CTM data in both summer and winter seasons, the UKF-based estimation is highly-correlated. For further comparison, the difference in NO<sub>2</sub> concentrations between OMI-retrieved data and TAPM-CTM values is considered for both the UKF-based estimation and original inventory. Fig. 3 shows the highest concentration of NO<sub>2</sub> is 24ppb in scale where as fig. 6a shows the highest concentration of NO<sub>2</sub> is 23ppb in scale. This difference is due to OMI-retrieved data. Similar trend shown in fig. 4 has the highest concentration of NO<sub>2</sub> is 13ppb in scale and fig. 6b has the highest concentration of NO<sub>2</sub> is 12ppb in scale. NO<sub>2</sub> emissions produced from combustion of fossil fuels, biomass burning, soil microbial activity and lighting. That is the main reason NO<sub>2</sub> concentration varies according to season. These type of variation is applicable to all seasons.

It should be noted that despite improving the estimation, the use of UKF and scaling factors could not fully eliminate the mismatch between TAPM-CTM and satellite-based emission concentrations. The reasons are likely with the use of constant scaling factors and conversion of the grid cells. Future work will therefore aim at finding not only an alternative to the current scaling approach but also a solution to combine the prediction model with satellite observations for better inversion of air-pollutant emissions.

## V. CONCLUSION

In this paper, we have presented a method to improve the emissions inventories using an unscented Kalman filter with a focus on estimation of power plant emissions in the GMR of New South Wales. Here, NO<sub>2</sub> concentrations estimated by TAPM-CTM are compared with emission levels retrieved by satellite. The results show generally a coincidence between the satellite data and the TAPM-CTM in which the UKF-based

estimation indicates a better match compared to the original inventory. The enhanced accuracy owing to the proposed UKF with a Matérn covariance function indicates an underestimation of the original inventory on the impact of power stations emissions, particularly in the winter. The proposed method assumes that the discrepancies in the modelled and observed NO<sub>2</sub> are due to the emissions estimated at the ground level. Nonlinearities and uncertainties associated with chemical processes, emissions aloft from lighting sources, air planes, and meteorological factors also contribute to the difference in NO<sub>2</sub> concentrations between the predicted model and satellite observations. The UKF has a limitation in terms of computation. A traditional Kalman filter allows retaining the moments of order 1 (mean) and 2 (covariance matrix). The Unscented Transform (UT) allows better conserving higher order moments of the states. In the UKF the sigma point matrix is recomputed after each propagation or measurement step from the covariance matrix, which only contains 2<sup>nd</sup> order moments, this higher order information is lost.

## ACKNOWLEDGMENT

This work is supported by University of Technology Sydney in a 2016 Data Arena grant and post-thesis publication scholarship. The authors would like to gratefully acknowledge the New South Wales Office of Environment and Heritage for providing data and laboratory facilities used in this study.

## REFERENCES

- [1] K. Berhane, C.-C. Chang, R. McConnell, W. J. Gauderman, E. Avol, E. Rapaport, R. Urman, F. Lurmann, and F. Gilliland, "Association of changes in air quality with bronchitic symptoms in children in California, 1993-2012," *JAMA*, vol. 315, no. 14, pp. 1491–1501, 2016.
- [2] A. K. Gorai, P. B. Tchounwou, and G. Mitra, "Spatial variation of ground level ozone concentrations and its health impacts in an urban area in India," *Aerosol and Air Quality Research*, vol. 17, no. 4, pp. 951–964, 2017.
- [3] J. P. Veefkind, J. F. de Haan, E. J. Brinksma, M. Kroon, and P. F. Levelt, "Total ozone from the ozone monitoring instrument (OMI) using the DOAS technique," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 5, pp. 1239–1244, May 2006.
- [4] W. Wang, L. Flynn, X. Zhang, Y. Wang, Y. Wang, F. Jiang, Y. Zhang, F. Huang, X. Li, R. Liu, Z. Zheng, W. Yu, and G. Liu, "Cross-Calibration of the Total Ozone Unit (TOU) With the Ozone Monitoring Instrument (OMI) and SBUV/2 for Environmental Applications," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 12, pp. 4943–4955, Dec 2012.
- [5] A. J. Krueger, N. A. Krotkov, K. Yang, S. Carn, G. Vicente, and W. Schroeder, "Applications of satellite-based sulfur dioxide monitoring," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 2, no. 4, pp. 293–298, Dec 2009.
- [6] E. J. Bucsela, E. A. Celarier, M. O. Wenig, J. F. Gleason, J. P. Veefkind, K. F. Boersma, and E. J. Brinksma, "Algorithm for NO<sub>2</sub> vertical column retrieval from the ozone monitoring instrument," *IEEE Trans. Geoscience and Remote Sensing*, vol. 44, no. 5, pp. 1245–1258, 2006.
- [7] B. Mijling and R. J. van der A, "Using daily satellite observations to estimate emissions of short-lived air pollutants on a mesoscopic scale," *Journal of Geophysical Research: Atmospheres*, vol. 117, no. D17, pp. n/a–n/a, 2012, d17302.
- [8] G. W. Leppelmeier, O. Aulamo, S. Hassinen, A. Malkki, T. Riihisaari, R. Tajakka, J. Tamminen, and A. Tanskanen, "OMI very fast delivery and the Sodankylä Satellite Data Centre," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 5, pp. 1283–1287, May 2006.
- [9] M. F. G. Ferreyra, G. Curci, and M. Lanfri, "First Implementation of the WRF-CHIMERE-EDGAR Modeling System Over Argentina," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 12, pp. 5304–5314, Dec 2016.

- [10] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 2, pp. 354–379, April 2012.
- [11] J. Xu, O. Schüssler, D. G. L. Rodriguez, F. Romahn, and A. Doicu, "A Novel Ozone Profile Shape Retrieval Using Full-Physics Inverse Learning Machine (FP-ILM)," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 12, pp. 5442–5457, 2017.
- [12] Z. Zeng, L. Lei, S. Hou, F. Ru, X. Guan, and B. Zhang, "A Regional Gap-Filling Method Based on Spatiotemporal Variogram Model of CO<sub>2</sub> Columns," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 6, pp. 3594–3603, June 2014.
- [13] S. Hogue, E. Marland, R. J. Andres, G. Marland, and D. Woodard, "Uncertainty in gridded CO<sub>2</sub> emissions estimates," *Earth's Future*, vol. 4, no. 5, pp. 225–239, 2016.
- [14] J. Ou, X. Liu, X. Li, and X. Shi, "Mapping Global Fossil Fuel Combustion CO<sub>2</sub> Emissions at High Resolution by Integrating Nightlight, Population Density, and Traffic Network Data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 4, pp. 1674–1684, April 2016.
- [15] A. L. Notte, S. Tonin, and G. Lucaroni, "Assessing direct and indirect emissions of greenhouse gases in road transportation, taking into account the role of uncertainty in the emissions inventory," *Environmental Impact Assessment Review*, vol. 69, pp. 82 – 93, 2018.
- [16] S. Monni, S. Syri, and I. Savolainen, "Uncertainties in the finnish greenhouse gas emission inventory," *Environmental Science & Policy*, vol. 7, no. 2, pp. 87 – 98, 2004.
- [17] M. Cope and S. Lee. (2009) Chemical transport model user manual. [Online]. Available: <https://www.cmar.csiro.au>
- [18] S. Li, L. Chen, X. Xiong, J. Tao, L. Su, D. Han, and Y. Liu, "Retrieval of the haze optical thickness in north china plain using modis data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 5, pp. 2528–2540, May 2013.
- [19] Z. Sun, J. Davis, and W. Gao, "Estimating Error Covariance and Correlation Region in UV Irradiance Data Fusion by Combining TOMS-OMI and UVMRP Ground Observations," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 1, pp. 355–370, Jan 2018.
- [20] A. B. Gilliland, R. L. Dennis, S. J. Roselle, and T. E. Pierce, "Seasonal NH<sub>3</sub> emission estimates for the eastern United States based on ammonium wet concentrations and an inverse modeling method," *Journal of Geophysical Research: Atmospheres*, vol. 108, no. D15, 2003, 4477.
- [21] S. Metia, S. D. Oduro, H. N. Duc, and Q. Ha, "Inverse air-pollutant emission and prediction using extended fractional kalman filtering," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 5, pp. 2051–2063, May 2016.
- [22] M. Azzi, H. Duc, and Q. Ha, "Towards Sustainable Energy Usage in the Power Generation and Construction Sectors - a Case Study of Australia," *Automation in Construction*, vol. 59, pp. 122–127, 2015.
- [23] S. Julier, J. Uhlmann, and H. Durrant-Whyte, "A new method for the nonlinear transformation of means and covariances in filters and estimators," *IEEE Transactions on Automatic Control*, vol. 45, no. 3, pp. 477–482, Mar 2000.
- [24] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2006.
- [25] (2016) Air Quality Management: Online Portal. <http://www.epa.gov>. [Online]. Available: <http://www.epa.gov/>
- [26] Healthy Environment, Healthy Community, Healthy Business. <http://www.epa.nsw.gov.au/air/airinventory.htm>.
- [27] H. L. V. Trees, K. L. Bell, and Z. Tian, *Detection Estimation and Modulation Theory, Part I: Detection, Estimation, and Filtering Theory*, second edition ed. John Wiley & Sons, New York, 2013.
- [28] S. Särkkä, "Unscented rauch-tung-striebel smoother," *Automatic Control, IEEE Transactions on*, vol. 53, no. 3, pp. 845–849, April 2008.
- [29] E. A. Wan and R. van der Merwe, *The Unscented Kalman Filter*. John Wiley & Sons, Inc., 2002, ch. 7, pp. 221–280.
- [30] S. Särkkä, "On unscented kalman filtering for state estimation of continuous-time nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 52, no. 9, pp. 1631–1641, Sept 2007.
- [31] Stay upto date. <http://www.environment.nsw.gov.au>.
- [32] S. López-Aparicio, M. Guevara, P. Thunis, K. Cuvelier, and L. Tarrasón, "Assessment of discrepancies between bottom-up and regional emission inventories in norwegian urban areas," *Atmospheric Environment*, vol. 154, pp. 285 – 296, 2017.
- [33] P. Levelt, G. van den Oord, M. Dobber, A. Malkki, H. Visser, J. de Vries, P. Stammes, J. Lundell, and H. Saari, "The ozone monitoring instrument," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 5, pp. 1093–1101, May 2006.
- [34] P. Addabbo, M. di Bisceglie, and C. Galdi, "The unmixing of atmospheric trace gases from hyperspectral satellite data," *IEEE Trans. Geoscience and Remote Sensing*, vol. 50, no. 1, pp. 320–329, Jan 2012.
- [35] M. J. Bechle, D. B. Millet, and J. D. Marshall, "Remote sensing of exposure to NO<sub>2</sub>: Satellite versus ground-based measurement in a large urban area," *Atmospheric Environment*, vol. 69, pp. 345 – 353, 2013.
- [36] S. Kembball-Cook, G. Yarwood, J. Johnson, B. Dornblaser, and M. Estes, "Evaluating NOx emission inventories for regulatory air quality modeling using satellite and air quality model data," *Atmospheric Environment*, vol. 117, pp. 1 – 8, 2015.
- [37] K. M. Han, S. Lee, L. S. Chang, and C. H. Song, "A comparison study between CMAQ-simulated and OMI-retrieved NO<sub>2</sub> columns over East Asia for evaluation of NO<sub>x</sub> emission fluxes of INTEX-B, CAPSS, and REAS inventories," *Atmospheric Chemistry and Physics*, vol. 15, no. 4, pp. 1913–1938, 2015.