Development of an infrared pollution index to identify ground-level compositional, particle size, and humidity changes using Himawari-8

M. Sowden
*Edith Cowan University*

D. Blake

D. Cohen

A. Atanacio

Ute Mueller
*Edith Cowan University*

Follow this and additional works at: https://ro.ecu.edu.au/ecuworkspost2013

Part of the *Oceanography and Atmospheric Sciences and Meteorology Commons*


This Journal Article is posted at Research Online.
https://ro.ecu.edu.au/ecuworkspost2013/8065
Development of an infrared pollution index to identify ground-level compositional, particle size, and humidity changes using Himawari-8

M. Sowdena, D. Blakeb, D Cohenc, A Atanacio, U. Muellerb

a Corresponding author, Miles Sowden, m.sowden@ecu.edu.au (ECU) School of Science, Edith Cowan University (ECU), Joondalup, Western Australia, 270 Joondalup Dr, Joondalup, Western Australia 6027, Australia

b Ute Mueller u.mueller@ecu.edu.au, David Blake d.blake@ecu.edu.au Centre for Ecosystem Management, School of Science, Edith Cowan University, Joondalup, Western Australia, Australia

c David Cohen dcz@ansto.gov.au, Armand Atanacio aaz@ansto.gov.au Centre for Accelerator Science, ANSTO, New Illawarra Rd, Lucas Heights, NSW, 2234, Australia

Highlights

• All 10 Himawari-8 infrared wavelengths used with 5 BTD indices;
• 3-year analysis of NO, NO2, SO2, O3, PM2.5, PM10, and aerosol species;
• Single pollutant events (meteorology) minimizes spectral contamination;
• Dynamic pollutant index derived from BTD indices;
• Composition, concentration, size and humidity derived from 5 BTD indices.

Abstract

Speciated air quality data informs health studies and quantitates impacts. However, monitoring is concentrated around populated regions whilst, large remote and rural regions remain unmonitored despite risks of dust-storms or wild-fires. Sub-hourly, infrared, geostationary data, such as the ten-minute data from Himawari 8, could potentially be used to quantify regional air quality continually. Monitoring of Aerosol Optical Depth (AOD) is restricted to visible spectra (i.e. daytime only), while newer quantification methods using geostationary infrared (IR) data have focused on detecting the presence, or absence, of an event. Limited attention has been given to the determination of particle size and aerosol composition (such as sulfates, black carbon, sea-salt, and mineral dust), using IR exclusively, and more appropriate methods are required to improve the understanding of source impacts.
Hourly data were collected for a three-year study period (July 2015 to July 2018) across the greater Sydney region in Eastern Australia from seventeen ground-based sites that measured meteorological data and quantified ambient concentrations of NO, NO₂, SO₂, PM₂.₅, PM₁₀, and O₃. This data was combined with source-apportioned categories (soil, sea-spray, smoke, secondary sulfates, and vehicles) from positive matrix factorization (PMF) of elemental aerosol collected on daily filters at five monitoring sites across the region. Regression analysis of five brightness temperature difference (BTD) infrared indices were used to determine a pollution index.

The pollution index was shown to be related to humidity, particle size, and compositional changes. Unlike fixed thresholds, the continual index function can be aggregated spatially and temporarily. PM₂.₅ and O₃ may be distinguished from each other based on spectral responses. However, BTD appears insensitive to concentration. The pollution index was useful for identifying composition prior to determining concentration.

**Keywords**: aerosol; Himawari-8; infrared; ground-level concentration

### 1. Introduction

The World Health Organization (WHO) estimates that globally 12.6 million deaths each year are attributable to unhealthy environments. Air pollution, both indoor (4.26 million deaths) and outdoor (3.73 million deaths) is a significant environmental factor in these deaths (Prüss-Üstün and Corvalán, 2006; WHO, 2016). The National Environment Protection Measure for Ambient Air (Air NEPM) sets national standards for six criteria air pollutants to which most Australians are exposed: carbon monoxide, ozone, sulfur dioxide, nitrogen dioxide, lead, and particulates (Australia Government, 2016).

Particulate matter is a significant criteria pollutant as it is ubiquitous and occurs at high concentration levels (Bennett et al., 2019). Particulate matter (PM) health effects include respiratory, lung cancer, and adverse cardiovascular effects (Barnett et al., 2012; Bell and Adams, 2008; Lin et al., 2016; Simpson et al., 2005; Tonne et al., 2016; Watanabe et al., 2016). There is some debate about the significance of size fractions of PM and whether specific sources (i.e. composition) of air pollution are more important than others (Bell et al., 2009; Goudie, 2014; Pirani et al., 2015). Urbanization, coupled with increased affluence, leads to an increase in vehicles and industrial activities with
concurrent increases in ambient particulate levels (Font et al., 2014; Kumar et al., 2014). Additionally, geogenic emissions from fires (Zielinski et al., 2016) and wind-blown events (Acosta-Martínez et al., 2015) contribute to the ubiquity of PM events in populated, arid, and forested regions. The composition (and indirectly health impacts) of air pollution are dependent on the source, e.g. desert dust is principally crustal but may include biological material (Goudie, 2014) while smoke from wildfires is chiefly organic, but may include inorganic compounds (Bell and Adams, 2008).

1.1. Surface monitoring requirements

Various methods have been used to estimate, predict, or measure airborne contaminant concentrations in order to quantify the population exposure (Csavina et al., 2012). Ambient monitoring using fixed locations provides reliable concentration data, often spanning decades. However, monitoring sites may be sparsely distributed in regional areas because of cost and infrastructure (power, cooling, and security) limitations (Melles et al., 2011; Wang et al., 2013) and situated according to population, rather than areas of risk (NSW, 2019).

Methods to determine concentration must have sufficient spatial and temporal resolution to accurately represent an event (typically minutes to hours or days) (Marshall et al., 2008; Zhao et al., 2016). The optimum monitoring interval is on the cusp of the turbulent zone, i.e. five to ten minutes, for air quality events, to achieve reproducible results and simultaneously be fast enough to detect those events (Sowden et al., 2019a). Despite this temporal requirement, the accredited PM sampling method (Chow, 1995, 1998) draws a predetermined volume of air over a daily period through a size-selective inlet and captures the total mass on a filter, which reduces short-term incident information to a smoothed daily average.

In contrast, electronic monitors, such as Beta attenuation monitors, record hourly or sub-hourly data (Bencs et al., 2010). Health-related studies suggest that the minimum requirements for monitoring atmospheric events are a spatial resolution of one kilometer and a temporal resolution of an hour (Chow, 1995, 1998). However, airborne concentrations may be below detection limits, even using twenty-four-hour filter samples (Harper, 2015). Sophisticated analysis techniques such as accelerator-based
ion beam analysis techniques (Cohen et al., 2014) may be required to detect pollutants at those low concentration levels on small microgram samples.

1.2. Accuracy of estimations

In areas without dedicated monitoring, it may be possible to estimate concentrations using dispersion modeling (Ancona et al., 2015) or remote sensing (Kloog et al., 2013) and infer compositions based on the source (Holmes and Morawska, 2006; Hopke, 2016; You et al., 2016b). Dispersion modeling takes the estimated emission from each source and disperses the pollutants according to parameterizations of the prevailing meteorology (Hewson, 1956; Stull, 1984).

For a simple source in flat terrain, dispersion modeling is very accurate. The US-EPA indicates the error margin to be a factor of two using period summations (maximum hourly daily and annual averages) and quantile plots (Wilson and Venkatram, 1998). Validation studies have identified where models successfully replicated annual averages and maxima but failed on paired hourly studies (Hurley et al., 2005). Most of the predicted inaccuracies lie with the input data assumptions (Sowden et al., 2008). When including a large domain, with multiple natural sources and summing small wind field errors over hundreds of kilometers, the errors rapidly increase and make dispersion modeling unsuitable over large regions (Draxler et al., 2015).

1.3. Remote sensing of ground-level concentrations

Remote sensing has been used to determine aerosol concentrations (Li et al., 2015; van Donkelaar et al., 2015; Wu et al., 2016; You et al., 2016a), where direct monitoring is unavailable (Li et al., 2016; Lin et al., 2015; Tsay et al., 2016). Aerosol Optical Depth (AOD) (Hsu et al., 2013; Levy et al., 2013) is a measure of the extinction of electromagnetic radiation by dust and haze, which can absorb or scatter light (ESRL, 2018). AOD is dimensionless and is related to the total amount of aerosol in the vertical column of the atmosphere over a location. AOD measurements have been extrapolated to surface concentrations (Achad et al., 2013; Bilal et al., 2016; Hyder et al., 2014; Kumar et al., 2007; Le et al., 2014; Muhammad and Nguyen Thi, 2015; Yoon et al., 2016; Zhang et al., 2012).

Short-term events such as fires, inversion weather conditions, and clouds should be taken into consideration as these events may be missed during infrequent satellite
overpasses (Baldassarre et al., 2015; Freeborn et al., 2014; O’Loingsigh et al., 2015; Philip et al., 2016; Zhang et al., 2011). Similarly, there is a need to extrapolate the total AOD column amount into respirable ground level concentrations (GLCs) to ensure comparative measurements (Bukowiecki et al., 2016; Sotoudeheian and Arhami, 2014). While reviews (Mhawish et al., 2018) and studies (Xu et al., 2015) both indicate high correlations (~ 0.8) for long term studies between AOD and ground-based measurements, paired daily studies report lower correlations which were attributed to the inter-day fluctuations of particle size (Bennouna et al., 2016).

1.4. Infrared indices

Remote sensing using both visible (scattering) and infrared (IR) wavelengths (detecting radiation) (She et al., 2018) has been used to monitor GLCs. Clouds impact reflectance (Levy et al., 2013) and the infrared portion of the electromagnetic spectrum (Schmit et al., 2018; She et al., 2018). While traditional methods for quantifying aerosol relies on using the daytime-only visible portion of the electromagnetic spectrum (Levy et al., 2013), more recent studies, favor revising and utilizing older infrared dust detection methodologies (Ackerman, 1997; Park et al., 2014) to provide continuous monitoring (i.e. 24 hours a day) (She et al., 2018).

Satellite algorithms supply the cumulative infrared absorption (from land, sea, water vapor (clouds), and trace atmospheric gases) as a brightness temperature (BT) measurement. Brightness temperature difference (BTD) between two infrared wavelengths (λ) (i.e. $\text{BTD}_{\lambda_1-\lambda_2 \, \mu m} = \text{BT}_{\lambda_1} - \text{BT}_{\lambda_2}$) (Shang et al., 2019; She et al., 2018; Sowden et al., 2018) is a simple high-pass filtering method that enhances the differences (trace gases) between two similar wavelength bands while minimizing commonality (land/sea and cloud) between the two wavelengths. Following the methodology of our earlier work (Sowden et al., 2019b), we utilized five brightness temperature difference (BTD) indices ($\text{BTD}_{3.9-6.2 \, \mu m}$, $\text{BTD}_{11-12 \, \mu m}$, $\text{BTD}_{9.6-13 \, \mu m}$, $\text{BTD}_{8.6-10 \, \mu m}$, and $\text{BTD}_{6.9-7.3 \, \mu m}$). These band combinations were chosen based on reducing correlations between indices and utilizing all ten IR wavelengths.

1.5. Aims of the study

This study seeks to understand if air quality events could be detected against background levels, identify which pollutants and source types contributed to the...
enhanced concentrations, and *quantitate* the concentrations of the events using the five, previously selected, BTD indices.

### 2. Data and study area

Three years (July 2015-2018) of surface concentration data were obtained from the New South Wales (NSW) Office of Environment and Heritage (OEH) for seventeen real-time monitoring sites across the greater Sydney region in eastern Australia. Figure 1 depicts the locations of these monitoring sites. Data from the Australian Nuclear Science and Technology Organization (ANSTO), for five co-located sites (blue squares in Figure 1) over the same monitoring period as the OEH data, were obtained. The ANSTO data were derived from sample analysis, which captured airborne PM$_{2.5}$ on filters over twenty-four hours twice weekly. These two concentration datasets were paired in time with ten-minute spectral data from the Himawari-8 geostationary satellite.

![Map of Sydney region with monitoring sites](map.png)
2.1. NSW Office of Environment and Heritage

Meteorological and ambient air quality concentration data was downloaded from the New South Wales (NSW) government office of Environment and Heritage (OEH) website https://www.environment.nsw.gov.au/AQMS/search.htm for the three years, July 2015 to 2018. An air pollution index (API) was calculated as the sum over all pollutants of the hourly data divided by the daily criteria level specified by the National Environment Protection Measure for Ambient Air (NEPM) standard. The 97.5 percentile of the data was used to approximate the daily criteria for pollutants that did not have a daily standard. This approach differed from OEH’s Air Quality Index (AQI), which used inconsistent timescales for each pollutant (hours for SO$_2$, NO, and NO$_2$ but daily for particulate matter). Accordingly, we have renamed our index as Air Pollution Index (API) to highlight the difference.

2.2. ANSTO data

The Australian Nuclear Science and Technology Organization (ANSTO) has applied accelerator-based nuclear techniques to the characterization of fine PM$_{2.5}$ quantification of ambient air as part of the Sydney Fine Particle Study (Cohen et al., 2016) and ongoing monitoring. Fine particles were collected on filters over twenty-four hours twice a week. Each of these filters was analyzed for 23 elemental and chemical species: hydrogen (H), sodium (Na), aluminum (Al), silicon (Si), phosphorus (P), sulfur (S), chlorine (Cl), potassium (K), calcium (Ca), titanium (Ti), vanadium (V), chromium (Cr), manganese (Mn), iron (Fe), cobalt (Co), nickel (Ni), copper (Cu), zinc (Zn), selenium (Se), bromium (Br), lead (Pb), black carbon (BC) and total nitrogen (Total N).

Positive matrix factorization (PMF) source apportionment techniques were applied to this data to identify different source components or fingerprints that make up the measured total PM$_{2.5}$ mass at each of these monitoring sites (Cohen et al., 2012; Cohen et al., 2014). Table 1 presents a summary of three years PMF data provided by ANSTO (July 2015 to 2018) for five sites (Liverpool, Mayfield, Musslebrook, Richmond, and Warrawong) across the greater Sydney region.

Table 1: Summary of three years PMF data provided by ANSTO (July 2015 to 2018) for five sites (Liverpool, Mayfield, Musslebrook, Richmond, and Warrawong) across the greater Sydney region.
ANSTO determined seven or eight characteristic fingerprints for each of the five ANSTO sites using PMF analysis. These fingerprints included soil, sea-spray, secondary sulfates (coal-fired power stations, oil refineries, motor vehicles, and industry), smoke (biomass burning, domestic wood heaters, and motor vehicles) and Auto1 (primary automobile source) components. Industrial emission profiles (Ca, S, Fe, and metals) and Auto2 (secondary automobile source) were only supplied where those profiles contributed significantly to a site.
Table 1: Average PM$_{2.5}$ concentration and aerosol percentages (by mass) PMF categories for each ANSTO site.

<table>
<thead>
<tr>
<th>Site</th>
<th>Liverpool</th>
<th>Mayfield</th>
<th>Musslebrook</th>
<th>Richmond</th>
<th>Warrawong</th>
</tr>
</thead>
<tbody>
<tr>
<td>PM$_{2.5}$ ($\mu$g/m$^3$)$^1$</td>
<td>8.37</td>
<td>6.98</td>
<td>5.84</td>
<td>6.89</td>
<td>9.47</td>
</tr>
<tr>
<td>Smoke</td>
<td>31.7%</td>
<td>26.2%</td>
<td>17.4%</td>
<td>35.1%</td>
<td>1.58%</td>
</tr>
<tr>
<td>Auto1</td>
<td>20.5%</td>
<td>23.8%</td>
<td>28.2%</td>
<td>11.2%</td>
<td>2.12%</td>
</tr>
<tr>
<td>Secondary Sulfate$^3$</td>
<td>17.1%</td>
<td>18.6%</td>
<td>21.4%</td>
<td>18.1%</td>
<td>16.9%</td>
</tr>
<tr>
<td>Industrial Sulfur</td>
<td>13.2%</td>
<td>NS</td>
<td>12.9%</td>
<td>18.0%</td>
<td>18.4%</td>
</tr>
<tr>
<td>Sea-spray$^3$</td>
<td>9.7%</td>
<td>17.0%</td>
<td>4.4%</td>
<td>5.6%</td>
<td>30.9%</td>
</tr>
<tr>
<td>Soil</td>
<td>3.22%</td>
<td>4.93%</td>
<td>6.55%</td>
<td>3.84%</td>
<td>12.42%</td>
</tr>
<tr>
<td>Industrial Ca</td>
<td>NS$^2$</td>
<td>1.87%</td>
<td>NS</td>
<td>NS</td>
<td>7.43%</td>
</tr>
<tr>
<td>Industrial Fe</td>
<td>NS</td>
<td>1.98%</td>
<td>NS</td>
<td>3.37%</td>
<td>4.81%</td>
</tr>
<tr>
<td>Industrial Metals</td>
<td>NS</td>
<td>NS</td>
<td>5.36%</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Auto2</td>
<td>1.39%</td>
<td>0.59%</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
</tbody>
</table>

1: Average hourly PM$_{2.5}$ concentration ($\mu$g/m$^3$).
2: NS indicates that the PMF parameter was not supplied for that site
3: Secondary Sulfate and sea-spray is abbreviated to SecS and sea respectively in subsequent discussions, and the Industrial metals are amalgamated into Industry (Ind)

2.3. Himawari-8

Raw data files of all ten infrared bands from the Himawari-8 satellite were obtained from the Australian Bureau of Meteorology (BOM) for the period from July 2015 to July 2018. These data files were cropped to the study domain and analyzed using the “Climate Data Operators” (CDO) (CDO, 2018). The raw brightness temperature data were averaged per hour (per pixel) and across the study domain (to obtain a temporal background absorbance which took into consideration diurnal temperature and atmospheric changes), and this hourly average was subtracted from the raw brightness temperature (per pixel) before calculating the BTD index value. Unlike the AOD methodology, the data was not spatially aggregated, nor was it screened based on cloud cover. The analysis considered five hourly BTD indices, which were matched in time to the hourly monitored OEH and interpolated hourly ANSTO data.

3. Methodology

3.1. Overview

It was assumed that concentration by event type was related to the five pre-selected BTD indices. However, to quantitate the concentrations, the composition, humidity, and particle size must be qualified, as these parameters influence the spectral response (Ackerman, 1997). This study analyzed the concentration distribution of different
pollution events over the three-year data period. A pollution index was developed, from
the BTD indices, which considered compositional, humidity, and particle size changes.
Predictive pollutant equations for relative humidity and percentage fine particulate
matter were quantified using regression from the pollution index and verified against
the two datasets, and a Weibull function was determined and used to estimate
compositional probability against the pollution index.

3.2. Variations in the measured OEH and ANSTO data

The variance of the data was investigated by classifying the measured concentration,
according to percentiles. The daily guideline for PM$_{2.5}$, PM$_{10}$, and SO$_2$ coincided with the
97.5 percentile, and this statistic was used for pollutants that did not have a daily
guideline. An Air Pollution Index (API) was defined as:

$$API = \sum_i \frac{\text{Concentration}_i}{\text{daily guideline}_i}$$  (1)

An air quality event (or incident) was indicated when at least one of the measured
pollutants exceeded the guideline. Carbon monoxide was subsequently removed from
the list of pollutants as it had few events compounded by a limited number of monitoring
sites. Competing spectral interferences from other pollutants were reduced by
restricting the analysis to single pollutant events. Using an air pollution index allowed
comparisons of severity between pollutants to be undertaken and enabled the temporal
addition of combined impacts from all pollutants to be determined.

Primary (emitted from source) and secondary (chemical reactions) aerosols are expected
from combustion by-products and photochemical reactions (Carter, 2007; Yarwood et
al., 2007) from both gaseous pollutants (NO, NO$_2$, SO$_2$, and O$_3$) and particulate matter
(PM$_{2.5}$ and PM$_{10}$). The OEH data indicated that the measured PM$_{10}$ included coarse
particles with an aerodynamic diameter (i.e. particle size) of 2.5 µm to 10 µm as there
were instances in the data where the concentration of PM$_{10}$ was less than PM$_{2.5}$ - which
could not occur if PM$_{10}$ included PM$_{2.5}$. The percentage of fine particles (0-100%) in the
measured OEH data was therefore defined as:

$$\text{Fines} = 100 \frac{\text{PM}_{2.5}}{\text{PM}_{2.5} + \text{PM}_{10}}$$  (2)

The data were screened to exclude background concentrations (i.e. where API < 1).
Boxplots were generated to determine the impacts of the site location, relative humidity
Development of a pollution index

Preliminary regression analysis at calibrating concentration scale factors per event type from the five BTD index values resulted in low correlations (0.05 to 0.3). Investigating the analysis of variance (ANOVA) of the API showed that the within variance (of a narrow range (1/100th) of the BTD index) was significantly greater than the between variances. This variance was further noted in the boxplots that were produced (Figure 2) (API by event type, severity, humidity, and particle size) that depicted both the relationship between the variables and the variance of the data. Removal of outliers and discrepancies between scale factors (slope, intercept, and median prediction) did not improve the predicted API/BTD regression correlations.

The failure of traditional regression methods necessitated an alternative approach. The OEH data analysis suggested that the API was related to changing event types, humidity, and particle size consequently, the variance within these categories was hindering quantification. A pollution index was therefore proposed that encompassed composition, humidity, particle size, and compositional changes.

The production of the boxplots (Figure 2) by humidity and percentage fines showed a consistent ordering of pollutants with three pollutants showing minimal differences, three moderate differences, and one major difference. These differences were used to determine a Bayesian “prior” pollution index with an initial index change of five, ten,
and twenty for minimal, moderate, and large differences, respectively. A Bayesian approach was used to include prior information that cannot be easily included via classical methods and where the data set includes a significant number of outliers and non-normal distributions (Lee and Wagenmakers, 2013). Bayesian regression was applied iteratively to determine the pollution index from the prior (dependent variable) against the five BTD indices as input (independent variables). Two BTD indices were subsequently discarded based on the lack of significance to the predicted coefficients from the final stepwise iteration.

### 3.4. Calculating concentration from the pollution index

The measured concentrations, relative humidity, and particle sizes were grouped and aggregated according to the integer of the pollution index. The aggregation minimized high concentrations from severe events, that may not represent the concentration across the entire Himawari pixel (4 km$^2$). Regression was used to determine coefficients for relative humidity, and particle size from the pollution index and maximum likelihood probabilities were determined by fitting a Weibull function (eq 4) to the event (composition) probabilities across the pollution index.

$$
Event\ \text{probability} = a \cdot \left(\frac{\beta}{\alpha}\right) \cdot \left(\frac{PI}{\alpha}\right)^{\beta-1} \exp\left(-\left(\frac{PI}{\alpha}\right)\right)
$$

(4)

Where:

- $a$ is a scale parameter to convert the cumulative sum (=1) to a constant maximum of unity in the graph, and according to the 99.9 percentile.
- $\alpha$ is the Weibull scale parameter that affects the curve width.
- $\beta$ is the Weibull shape parameter that changes the position of the maxima where one is asymptotic at low PI values, and five is asymptotic at high PI values; and
- PI is the pollutant index.
4. Results

4.1. Variations in the measured OEH and ANSTO data

4.1.1. Percentiles

The data recovery rates in Table 2 reflected that not all pollutants were measured at each site. Table 2 shows that while the guidelines (97.5%) represented most of the measured data, they accounted for only seven percent of the measured OEH concentrations. The skewed data distribution is further reflected in the fivefold increase in concentrations between the 99.9 percentile and the maximum.

Table 2: Total data recovery rates (including parameters not monitored at sites) and concentration percentiles from a) the OEH dataset, and b) the ANSTO dataset. The number of valid measurements per element is provided with the values in brackets representing the percentage of valid data recovered from total data records for each element.

<table>
<thead>
<tr>
<th></th>
<th>a) OEH data</th>
<th>b) ANSTO data</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>SO\textsubscript{2} pphm\textsuperscript{1}</td>
<td>PM\textsubscript{2.5} μg/m\textsuperscript{3}</td>
</tr>
<tr>
<td></td>
<td>NO pphm</td>
<td>Smoke μg/m\textsuperscript{3}</td>
</tr>
<tr>
<td></td>
<td>NO\textsubscript{2} pphm</td>
<td>Sea μg/m\textsuperscript{3}</td>
</tr>
<tr>
<td></td>
<td>CO ppm</td>
<td>Auto1 μg/m\textsuperscript{3}</td>
</tr>
<tr>
<td></td>
<td>O\textsubscript{3} pphm</td>
<td>Soil μg/m\textsuperscript{3}</td>
</tr>
<tr>
<td>486014 (100%)</td>
<td>395504 (81%)</td>
<td>31293 (94%)</td>
</tr>
<tr>
<td></td>
<td>428122 (88%)</td>
<td>31317 (94%)</td>
</tr>
<tr>
<td></td>
<td>428530 (88%)</td>
<td>31297 (94%)</td>
</tr>
<tr>
<td></td>
<td>158650 (33%)</td>
<td>31325 (94%)</td>
</tr>
<tr>
<td></td>
<td>349422 (72%)</td>
<td>24583 (74%)</td>
</tr>
<tr>
<td></td>
<td>481059 (99%)</td>
<td>31373 (94%)</td>
</tr>
<tr>
<td></td>
<td>402973 (83%)</td>
<td>19461 (58%)</td>
</tr>
<tr>
<td></td>
<td>243973 (50%)</td>
<td></td>
</tr>
<tr>
<td>50%</td>
<td>0</td>
<td>5.8</td>
</tr>
<tr>
<td>95%</td>
<td>0.5</td>
<td>20.4</td>
</tr>
<tr>
<td>97.5%</td>
<td>0.7</td>
<td>25.9</td>
</tr>
<tr>
<td>99.0%</td>
<td>1.2</td>
<td>35.3</td>
</tr>
<tr>
<td>99.9%</td>
<td>3.4</td>
<td>61.2</td>
</tr>
<tr>
<td>100%</td>
<td>21</td>
<td>89.7</td>
</tr>
</tbody>
</table>

\textsuperscript{1}: pphm = Parts per hundred million

4.1.2. Quantifying impacts by site, relative humidity, and particle size

There were minimal differences in measured concentrations between locations despite urban, rural, coastal, and forested land-use differences. Most events had an API of three to five (Figure 2a-b) and were within the 95\textsuperscript{th} percentile of event concentrations but
PM$_{10}$, and to a lesser extent SO$_2$, had a considerable number of exceedances above the 95% percentile level.

NO events coincided with high relative humidity (Table 2 and Figure 2a), while NO, NO$_2$, SO$_2$, and PM$_{10}$ all reflected an inverse concentration relationship with relative humidity. PM$_{10}$ events were the only events that occurred significantly during low humidity (0 to 20%) potentially from enhanced windblown dust events over dry soil. Ozone and PM$_{2.5}$ showed a reversed trend with concentrations increasing with relative humidity and reached a maximum at about 80% humidity, and this can be ascribed to moisture coagulating on fine particles (and increasing their mass) with wet deposition potentially occurring above 80% humidity.

Figure 2b shows that coarse particles dominated PM$_{10}$ measurements (as expected by the definition of $> 2.5$ µm) and NO events, while PM$_{2.5}$ events were from fine particles with a concentration (API) that decreased according to particle size (arguably related to volume). Ozone concentrations also exhibit an inverse concentration relationship with particle size. Similar, but weaker, trends were noted in the ANSTO data (Figure 2c-d).

**Event types**

Table 3 lists the most frequent event types in descending order. From the table, it is apparent that air quality events in the OEH dataset were dominated (80%) by six single-pollutant incidents, five two-pollutant incidents (10%), and the remaining (10%) occurred from low-frequency, multi-component incidents. While there are significant differences in maximum API concentrations, each singular pollutant event type had an API concentration of approximately 3.0. While the lack of PMF categories across all sites precluded a similarly detailed analysis of the ANSTO dataset, seventy percent of the ANSTO events were related to a single PMF source category. Concentration maxima from smoke and sea-spray were double that of the other categories.
Figure 2: Box and whisker plots by: a) OEH relative humidity; b) OEH percentage fines; c) ANSTO relative humidity, d) ANSTO percentage fines depicting API per pollutant (where a singular pollutant exceeded an API of unity for each monitoring site.)
Table 3: The most commonly occurring event types for a) the OEH API dataset and b) the ANSTO PM\(_{2.5}\) dataset monitored over the three years (July 2015 to 2018) at the selected monitoring sites.

<table>
<thead>
<tr>
<th>Event Type</th>
<th>No of events</th>
<th>Percent</th>
<th>Cumulative Percent</th>
<th>Std. Deviation</th>
<th>Mean</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>a) OEH data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PM(_{10})</td>
<td>8635</td>
<td>18.2</td>
<td>18.2</td>
<td>0.8</td>
<td>2.9</td>
<td>16.2</td>
</tr>
<tr>
<td>O(_{3})</td>
<td>7033</td>
<td>14.8</td>
<td>33.1</td>
<td>0.6</td>
<td>2.5</td>
<td>5.5</td>
</tr>
<tr>
<td>SO(_{2})</td>
<td>6665</td>
<td>14.1</td>
<td>47.1</td>
<td>1</td>
<td>3.6</td>
<td>13.2</td>
</tr>
<tr>
<td>NO</td>
<td>5605</td>
<td>11.8</td>
<td>59</td>
<td>0.6</td>
<td>3.4</td>
<td>6.0</td>
</tr>
<tr>
<td>PM(_{2.5})</td>
<td>5324</td>
<td>11.2</td>
<td>70.2</td>
<td>0.6</td>
<td>2.8</td>
<td>5.5</td>
</tr>
<tr>
<td>NO(_{2})</td>
<td>4789</td>
<td>10.1</td>
<td>80.3</td>
<td>0.6</td>
<td>2.8</td>
<td>5.4</td>
</tr>
<tr>
<td>TSP (PM(<em>{2.5})+PM(</em>{10}))</td>
<td>1394</td>
<td>2.9</td>
<td>83.2</td>
<td>2.1</td>
<td>4.8</td>
<td>22.5</td>
</tr>
<tr>
<td>NO(<em>{x}) (NO+NO(</em>{2}))</td>
<td>1100</td>
<td>2.3</td>
<td>85.6</td>
<td>0.7</td>
<td>4.1</td>
<td>7.3</td>
</tr>
<tr>
<td>O(<em>{3})+PM(</em>{10})</td>
<td>804</td>
<td>1.7</td>
<td>87.3</td>
<td>0.6</td>
<td>3.6</td>
<td>7.0</td>
</tr>
<tr>
<td>NO+CO</td>
<td>599</td>
<td>1.3</td>
<td>88.5</td>
<td>0.7</td>
<td>4.2</td>
<td>6.7</td>
</tr>
<tr>
<td>NO+PM(_{2.5})</td>
<td>598</td>
<td>1.3</td>
<td>89.8</td>
<td>0.8</td>
<td>4.9</td>
<td>8.4</td>
</tr>
<tr>
<td>TSP+NO(_{x})+CO</td>
<td>571</td>
<td>1.2</td>
<td>91</td>
<td>1.7</td>
<td>6.6</td>
<td>16.6</td>
</tr>
<tr>
<td>NO(_{x})+CO</td>
<td>562</td>
<td>1.2</td>
<td>92.2</td>
<td>0.9</td>
<td>4.7</td>
<td>8.0</td>
</tr>
<tr>
<td>SO(<em>{2})+PM(</em>{10})</td>
<td>494</td>
<td>1</td>
<td>93.2</td>
<td>1.4</td>
<td>4.9</td>
<td>11.9</td>
</tr>
<tr>
<td>NO+PM(_{10})</td>
<td>383</td>
<td>0.8</td>
<td>94</td>
<td>0.8</td>
<td>4.1</td>
<td>8.4</td>
</tr>
<tr>
<td><strong>b) ANSTO data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SecS</td>
<td>411</td>
<td>13</td>
<td>13</td>
<td>4.6</td>
<td>16.8</td>
<td>36.4</td>
</tr>
<tr>
<td>Smoke</td>
<td>408</td>
<td>12.9</td>
<td>25.9</td>
<td>8.5</td>
<td>24.5</td>
<td>89.2</td>
</tr>
<tr>
<td>Sea</td>
<td>401</td>
<td>12.6</td>
<td>38.5</td>
<td>6.3</td>
<td>17.6</td>
<td>60.1</td>
</tr>
<tr>
<td>Auto1</td>
<td>325</td>
<td>10.3</td>
<td>48.8</td>
<td>4.3</td>
<td>17.9</td>
<td>37.5</td>
</tr>
<tr>
<td>Soil</td>
<td>196</td>
<td>6.2</td>
<td>55</td>
<td>4.6</td>
<td>13.3</td>
<td>27.4</td>
</tr>
<tr>
<td>S</td>
<td>194</td>
<td>6.1</td>
<td>61.1</td>
<td>4.8</td>
<td>18.8</td>
<td>35.5</td>
</tr>
<tr>
<td>Auto2</td>
<td>115</td>
<td>3.6</td>
<td>64.7</td>
<td>4.8</td>
<td>12.3</td>
<td>26.9</td>
</tr>
<tr>
<td>Fe</td>
<td>98</td>
<td>3.1</td>
<td>67.8</td>
<td>5.3</td>
<td>17.5</td>
<td>30.7</td>
</tr>
<tr>
<td>Metals</td>
<td>58</td>
<td>1.8</td>
<td>69.6</td>
<td>4.5</td>
<td>9.3</td>
<td>20.4</td>
</tr>
<tr>
<td>Smoke+Auto1</td>
<td>55</td>
<td>1.7</td>
<td>71.3</td>
<td>9.1</td>
<td>33.9</td>
<td>60.8</td>
</tr>
<tr>
<td>Ca</td>
<td>17</td>
<td>0.5</td>
<td>71.8</td>
<td>2.2</td>
<td>12.1</td>
<td>17.2</td>
</tr>
</tbody>
</table>

Comparing ANSTO and OEH data

The correlation between the ANSTO and OEH datasets was 0.760 for the hourly ANSTO interpolation and 0.625 for the daily OEH summation. This correlation is less than expected when comparing between two analytical methods and is more typical of comparisons between remote sensing AOD and ground-based AERONET correlations (Bennouna et al., 2016; Mhawish et al., 2018; Yumimoto et al., 2016). While the different timescales may contribute to the low correlation on the ANSTO interpolation, a higher correlation would be expected on the daily OEH summation if the timescales were the cause of the low correlation. The low correlation arises from location differences and indicates the plume’s spatial variability rather than compositional differences between the two datasets.
4.2. Development of a pollution index

4.2.1. API by BTD boxplots

Boxplots (Figure 3) were produced for each of the five BTD indices, per event type and severity (the integer of API), showed that half of the plots depict a lack of change in concentration against the BTD index value (e.g. NO and BTD_{3.9-6.2 \ \mu m}). The lack of sensitivity in these plots accounts for the high-within, low-between ANOVA results. The index values were similar across three pollutant categories: [NO, NO\textsubscript{2}, and PM\textsubscript{2.5}]; [PM\textsubscript{10}, and SO\textsubscript{2}]; and O\textsubscript{3} with minimal changes within a group.

Low background concentrations covered a broader spectral response in the data than the events due to the larger dataset (97.5%) and the combined spectral result of other the pollutants. In contrast, some extreme events had a low frequency of observance, and the lack of data is evident in the narrow spectral range, which did not follow the trend of the other data (e.g. NO for BTD_{11-12 \ \mu m}). These background and low-frequency events were excluded from subsequent analysis.

Nitric oxide (NO) has a low spectral sensitivity to the five BTD indices, with only BTD_{6.9-7.3 \ \mu m} showing a definitive relationship in the boxplots. Similarly, ozone is best qualified by the singular BTD_{9.6-13 \ \mu m} index, although both BTD_{3.9-6.2 \ \mu m} > 50\degree C and BTD_{11-12 \ \mu m} > 2\degree C are potentially useful as thresholds. Sulfur dioxide (SO\textsubscript{2}) has the best spectral sensitivity across all five BTD indices, followed by nitrogen dioxide (NO\textsubscript{2}) across three indices. Both PM\textsubscript{2.5} and PM\textsubscript{10} exhibit a lack of spectral sensitivity to concentration.
4.2.2. Determining the pollution index

Iterative regression analysis to predict the pollution index (PI) determined the following BTD coefficients.

\[
\text{Pollution index} = 17.3 + 4.69BTD_{14-15} + 0.499BTD_{9-10} + 0.577BTD_{12-16} \tag{5}
\]
Figure 4 presents boxplots depicting the variation of the pollutant index across event types, fine fraction, and relative humidity. The pollutant type has the most significant impact on the pollution index with NO, NO$_2$, and PM$_{2.5}$ having low index values, PM$_{10}$ and SO$_2$ having intermediate levels and O$_3$ the highest. There is an inverse relationship between humidity and pollutant index, while the particle size relationship is less well defined.
Figure 4: Pollution index (PI) as a function of a) OEH events by relative humidity, b) OEH events by percentage fine aerosol, c) ANSTO events by relative humidity, and d) ANSTO events by percentage fine aerosol.

4.3. Determining concentration from the pollution index

Figure 5 depicts the mean concentrations and probabilities aggregated across the integer of the spectral pollution indices. Figure 5a demonstrates high correlations between relative humidity and percentage fines with the pollution index (PI). The regression coefficients were scaled to span the data range and simplified to:

Relative humidity = 100 – 10 PI/6 \hspace{1cm} (6)

Fines = 100 – 6PI + PI^2/10 \hspace{1cm} (7)

Figure 5b depicts the probability that the predicted concentration is below the mean (P50) or above the event threshold (97.5%) according to the pollution index. At a pollution index of 14, there is a 60% probability that an hourly measurement would be below the annual mean. With a pollution index in the range of 8 to 20 (blue shaded region) the probability that the measurement is below the mean is greater than the uncorrelated 50th percentile.

Similarly, with a pollution index in the range of 8 to 30 (red shading), there is less of a probability that it is an event. At lower or higher pollution index values, the probability decreases that the hourly concentration is below the annual mean and simultaneously increases the probability of an event. This relationship is more robust at higher pollution index values, and at a pollution index of 40, there is an eighty percent probability that the measurement is above the mean and the probability that it is an event doubles from 2.5% (100-97.5%) to 5%.

Figure 5 (c and d) shows that for most pollutants, the measured concentration is independent of the pollution index. SO2 and O3 show a slight positive correlation against index values, and the ANSTO data reflects a cut-off at an index value of about 25 in accordance with the PM2.5 data in Figure 5c. The ANSTO data reflects a much smaller dataset (two days per week), which was interpolated to hourly estimates, and consequently, there are fewer exceedances and more “noise” in the data. Figure 5d has been simplified to include only three PMF components to minimize overlapping factors.
The probability of an event type is related to the pollutant index for the OEH data (Figure 5e), but less dependence was noted for the ANSTO data (Figure 5f). The probability of a PM$_{2.5}$ event increases hyperbolically with decreasing pollution index with a probability of 0.1 at an index value of 25 increasing to 0.5 at an index of 5. Ozone increases with increasing index values, in contrast to all other pollutants, such that at an index value of greater than thirty-five is most probably an ozone event. SO$_2$ and PM$_{10}$ are similar, with both having a quadratic function with a maximum at an index value of about 25. Both NO$_x$ components (NO and NO$_2$) have similar quadratic profiles with little probability differences between them and overlap the PM$_{2.5}$ probability profile except at low index values (<10) where PM$_{2.5}$ is asymptotic while NO$_x$ decreases to zero. The derived Weibull parameters are supplied in Table 4.

Table 4: Weibull parameters per pollutant/aerosol factor sorted according to increasing shape parameter ($\beta$)

<table>
<thead>
<tr>
<th>Factor</th>
<th>PM$_{2.5}$</th>
<th>Smoke</th>
<th>Auto</th>
<th>NO</th>
<th>NO$_2$</th>
<th>PM$_{10}$</th>
<th>S</th>
<th>SO$_2$</th>
<th>Soil</th>
<th>Sea</th>
<th>SecS</th>
<th>O$_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>1.03</td>
<td>1.27</td>
<td>1.68</td>
<td>1.88</td>
<td>1.96</td>
<td>2.23</td>
<td>2.84</td>
<td>2.93</td>
<td>3.19</td>
<td>3.21</td>
<td>3.41</td>
<td>4.68</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>17.6</td>
<td>13.2</td>
<td>18.5</td>
<td>15.4</td>
<td>20.4</td>
<td>31.1</td>
<td>29.9</td>
<td>36.3</td>
<td>30.9</td>
<td>27.0</td>
<td>45.2</td>
<td>48.8</td>
</tr>
<tr>
<td>Scale</td>
<td>1483</td>
<td>637</td>
<td>336</td>
<td>291</td>
<td>96</td>
<td>4645</td>
<td>376</td>
<td>107</td>
<td>296</td>
<td>453</td>
<td>430</td>
<td>202</td>
</tr>
</tbody>
</table>

Similarities in the probability distributions were noted between the OEH and ANSTO data specifically for the O$_3$-Secondary Sulfates; PM$_{2.5}$-Smoke; and NO$_x$-Auto pairs. This similarity is also seen in the comparisons of event types with relative humidity and percentage fines as depicted in Figure 2.
Development of a pollution index using IR from Himawari-8

Figure 5: Average concentration and probability as a function of the pollution index. a) OEH relative humidity and particle size; b) probability below the mean or above the guideline (97.5%); c/d) OEH/ANSTO concentration; e/f) OEH/ANSTO probability; g/h) OEH/ANSTO probability as a Weibull function.
5. Discussion

5.1. Variations in the measured OEH and ANSTO data

5.1.1. Percentiles

The skewed concentration data indicates that air quality events reflect the log-normal nature of environmental data. Quantifying air quality events in standard statistical terms of means and population variances describes these sample populations poorly, as illustrated by comparing the measured maximum concentration of PM$_{2.5}$ of 335 µg/m$^3$ against the mean of 6.5 µg/m$^3$.

The concentrations reflect the cumulative probabilities of the wind blowing towards a sensor, the homogeneity of the plume over the monitoring area (4 km$^2$ for Himawari), the probability of an emission - which may be transient (such as smoke from a fire, wind-blown dust, passing vehicles), and the probability of favorable dispersion conditions (stability and wind speed). Average background concentrations are two to three orders of magnitude lower than the maximum concentrations, and it is this large skewed distribution between maxima, incident, and background concentrations that make it challenging to describe concentrations in terms of means and standard deviations.

5.1.2. Quantifying impacts by site, relative humidity, and particle size

Minor differences were noted between site locations despite differences in land type (sea/urban/forest/rural) between the sites. Both particle size and relative humidity demonstrated evidence of a causal relationship with concentration across all pollutant types. The weaker trends in the ANSTO data were ascribed to less data, the PMF categorization, and from interpolating daily data to hourly.

5.1.3. Event types

The data shows that a single pollutant per incident dominates air quality events. This indicates that the different event types seldom impact a receptor simultaneously, and this assists the quantification as multi-spectral contamination is reduced. However, it also indicates the dynamic nature of air pollution as one event type may be subsiding while another may be initiating (e.g. vehicular activity may be diminishing at sunset, but simultaneously the land-sea temperature differentials may initiate a wind directional change bringing in salt or crustal material) and the resulting mix of
pollutants may, therefore, have continually changing compositional and particle size attributes.

5.1.4. Comparing ANSTO and OEH data
The correlation coefficient between the OEH and ANSTO data is less than expected between two analytical methods. This lower correlation may be due to: a) narrow plumes which may partly miss the other site; b) different monitoring methods (electronic or filter) which may respond differently to changes in relative humidity (as filters would be indicative of the final relative humidity and/or dried prior to analysis in contrast to the instantaneous amount of moisture adsorbed on the particles); and c) differences in the temporal resolution of the two sampling methods requires an interpolation from the daily ANSTO measurements equivalent to the hourly NSW EPA data. Nevertheless, the correlation coefficient between the two monitoring methods represents a “best-case” scenario that could be expected when comparing predictive methodologies with inherently higher uncertainty.

5.2. Pollution index

5.2.1. API by BTD boxplots
The BTD by event and severity boxplots (Figure 3) demonstrate that approximately half of the concentration relationships are insensitive to the BTD index. NO, NO\textsubscript{2}, and PM\textsubscript{2.5} exhibit similar spectral responses to changing concentration, as does PM\textsubscript{10} and SO\textsubscript{2}. There is some justification in determining thresholds for ozone (e.g. BTD\textsubscript{3.9-6.2 μm} > 50°C and BTD\textsubscript{11-12 μm} > 2°C as only BTD\textsubscript{9.6-13 μm} depicts a concentration index relationship. However, from the plots, there are no apparent thresholds for the other pollutants that would simplify qualification and would not be best served by setting upper and lower bounds as per the OEH data summary.

Based on the BTD by event and severity boxplots (Figure 3) ozone, sulfur dioxide, and nitrogen dioxide exhibit the best spectral sensitivity to changes in concentration while particulate matter (both PM\textsubscript{2.5} and PM\textsubscript{10}) are intermediate and nitric oxide has reduced spectral sensitivity to the BTD indices.

Determining the pollution index
Dispersion of an air pollution episode takes time to dissipate, and consequently, air pollution changes constantly in composition and magnitude. The advantage of using a pollution index rather than a classification by event type is that it allows for a gradual transitioning (i.e. a continual function rather than a binary true/false) between pollutants considering changes in event type, particle size, and humidity. Figure 4 depicts the gradual transition in the index as the dominance and emission intensity of sources change, or temperature and evaporation effects affect the aerosol composition. Aggregation of the pollutant index is possible, both spatially and temporally, due to the continual function, unlike categorical variables such as event types, which cannot be averaged between categories. Finally, the resultant equations are simpler to code than a series of nested if-statements per site-specific threshold.

The pollutant index is derived from six out of the ten infrared BT absorbances and utilizes three BTD components BTD\textsubscript{11-12} µm, BTD\textsubscript{9-10} µm, and BTD\textsubscript{9.6-13.3} µm. Two BTD indices were rejected BTD\textsubscript{3.9-6.2} µm and BTD\textsubscript{8.6-10} µm because of low significance to the regression coefficients (for this study). Using six wavelengths enables most of the spectral variance to be described by these three indices and enables a more comprehensive pollutant index to be developed than using only three or four wavelengths. BTD\textsubscript{11-12} µm has traditionally been used to detect AOD. BTD\textsubscript{9-10} µm is a strong predictor of water vapor and indicates the difference between high and low cloud, which affects the vertical dispersion of the plume. Finally, BTD\textsubscript{8.6-10} µm gives a measure of the ozone column. Therefore, these three indices can be expected to contribute to the determination of particle size, relative humidity, column to ground level, and an indicator of photochemical precursor species. While the pollution index methodology could be extended to other regions, the coefficients are expected to be different because of the location-specific thresholds (Shang et al., 2019), moisture, and particle sizes.

5.3. Calculating concentration from the pollution index

The frequency distribution (Table 1) and boxplots (Figure 2) depict the full range of concentrations that occur in the measured datasets. However, the average concentration is insensitive to the pollution index, as depicted in Figure 5(c & d), and this explains why BTD has been used to detect but not quantify events. It has previously been suggested (Ackerman, 1997) that the subtraction of the two-wavelength channels
accounts for the loss of sensitivity and that quantification should use the raw spectra instead (after correcting of atmospheric effects).

Fitting a Weibull function on the probability data enables an assessment of the resolving power of the pollution index (i.e. the ability of the pollution index to categorize the event type). Ozone and PM$_{2.5}$ are inversely correlated, and as one increases, the other decreases. Unfortunately, the other components are not well resolved, and for example, soil, sea-spray, and SO$_2$ are highly correlated and challenging to resolve. Additional external information, such as distance to coast or wind direction, may enable the resolution to be enhanced. Table 4 provides the Weibull parameters, which enable a coarse assessment of concentration to be determined, assuming that a probability of one equates to the 99.9 percentile.

The high correlation between the probability distributions of the OEH and ANSTO datasets were noted. It was rationalized that this was a causal relationship, and not merely auto-correlation based on the relative humidity and particle size because the ANSTO PMF factors do generate the OEH components. Smoke does generate PM$_{2.5}$; coal-fired power stations (the main contributor to secondary sulfates) are contributing factors in ozone formation; wind-blown soil and sea-spray does generate sulfates; and motor vehicles do generate NO$_x$. This is an interesting observation in that it shows that a) the PMF analysis of particulate matter on the filters can estimate the gaseous pollutant concentrations; b) it verifies the PMF analysis; and c) indicates that the gaseous surface concentrations can be detected using remote sensing infrared wavelengths.

6. Conclusion

Data relating to air quality events are best interpreted according to percentiles rather than means and population variances as the concentrations are dependent on the cumulative probabilities of wind direction, speed, and emissions factors. While air quality events are typically dominated by a single event type per incident, the dynamic nature of air pollution implies a constant change in composition, absorbed moisture, and particle sizes. Significant differences were noted between OEH and ANSTO’s PM$_{2.5}$ concentrations, and this was attributed to narrow plumes, different monitoring methods, and temporal resolution between the two sampling techniques. These
differences are expected to be amplified when comparing data across a coarse (4 km²) satellite monitoring grid.

The pollution index allows for a gradual temporal transitioning between pollutants, particle size, and humidity, and the continuous index scale can be aggregated both spatially and temporally. While the five BTD indices are related to chemical composition, and this is independent of location, the pollution index is dependent on moisture and particle size, and it is therefore expected to be location-specific. Weibull parameterization of the probability distributions allows an estimate of the probability of particulate matter source factors (ANSTO PMF) and gaseous events (OEH) to be determined. While some compounds are well resolved (e.g. PM$_{2.5}$ and O$_3$), others (e.g. PM$_{10}$ and SO$_2$) are poorly resolved. External information such as wind direction may assist in resolving the event type.

Unfortunately, BTD is unsuitable for quantification as the low regression correlations prevent a direct quantification of concentrations. However, the predicted probabilities can be scaled to the maxima (or average) concentration, and in so doing provide period statistics that are comparative to dispersion modeling in accuracy but with the benefit of no wind field errors (as the plume is where it is observed) and sporadic events are detected in contrast to dispersion modeling which cannot model unknown sources.

The pollution index is in the process of being validated both spatially and temporally (i.e. is a sea-spray plume coherent across space and does it develop and dissipate according to the prevailing meteorology) and verified for global applicability to investigate if it is susceptible to underlying local land use (e.g. sea, sand, and vegetation) effects. Preliminary investigations indicate that prior screening, based on cloud cover and atmospheric stability, may improve quantification and that the plumes are coherent across space and time (see https://www.linkedin.com/posts/miles-sowden-6a1b704_duststorm-remotesensing-aerosol-activity-6586901543742726144-aglZ).
7. References


composition and sources of fine atmospheric particles over six different areas of Northern Belgium. Environmental Pollution 158, 3421-3430.


quantification of coal-fired power station pollution in metropolitan Sydney. Atmospheric Environment 61, 204-211.


ESRL, Earth System Research Laboratory, Global Monitoring Division, 2018. SURFRAD Aerosol Optical Depth.


Development of a pollution index using IR from Himawari-8


Development of a pollution index using IR from Himawari-8


Wilson, R., Venkatram, A., 1998. Model evaluation results for AERMOD.


determined from the GOCI satellite instrument. Atmospheric Chemistry and Physics 15, 13133-13144.


8. Appendix

8.1. SPSS OEH syntax code

* Do some calculations.
USE ALL.
compute pSize=RND(100*PM25/(PM25+PM10)).
formats pSize (f3.0).
compute API = 0.00.
formats API (f5.1).
compute API = API + O3 / 4.3.
compute API = API + SO2 / 0.7.
compute API = API + NO / 5.8.
compute API = API + NO2 / 2.7.
compute API = API + PM25 / 25.
compute API = API + PM10 / 50.
EXECUTE.

* Categorize on integers to display in graphs.
compute severity=0.
formats severity (f3.0).
if (API >1) severity=rnd( (API)).
if severity > 9 severity=9.

* Do a boolean test on the daily criteria (or 97.5 if no daily), ordered by increasing N.
compute event=0.
formats event (f3.0).
if (CO > 0.7) event = 1.
if (O3 > 4.3) event = event + 2.
if (SO2 > 0.8) event = event + 4.
if (NO > 5.8) event = event + 8.
if (NO2 > 2.7) event = event + 16.
if (PM25 > 25) event = event + 32.
if (PM10 > 50) event = event + 64.

* Give a text label to the classification.
string etype(a10).
compute etype=STRING(event,N3).
compute etype="other".
  * Singular events.
  if (event=0) etype = "low".
  if (event=64) etype = "PM10".
  if (event=8) etype = "NO".
  if (event=4) etype = "SO2".
  if (event=2) etype = "O3".
  if (event=16) etype = "NO2".
  if (event=32) etype = "PM2.5".
  if (event=1) etype = "CO".
  * Severe events.
  if (event=96) etype = "TSP".
  if (event=121) etype = "TSP+NOx+CO".
  if (event=89) etype = "TSP+NOx+CO".
  if (event=97) etype = "TSP+CO".
  if (event=113) etype = "TSP+NOx+CO".
  if (event=105) etype = "TSP+NOx+CO".
  if (event=112) etype = "TSP+NO".
  if (event=98) etype = "TSP+O3".
  if (event=57) etype = "TSP+NOx+CO".
  * High events.
if (event=24) etype = "NOx".
if (event=25) etype = "NOx+CO".
if (event=9) etype = "NO+CO".
if (event=66) etype = "O3+PM10".
if (event=41) etype = "NO+PM2.5+CO".
if (event=80) etype = "NO+PM10".
* low events.
if (event=40) etype = "NO+PM2.5".
if (event=68) etype = "SO2+PM10".
if (event=80) etype = "NO+PM10".
* classify BTD into ~ 100 categories.
compute I7_8=rnd(B7_8).
compute I14_15=rnd(8*B14_15)/8.
compute I9_10=rnd(6*B9_10)/6.
compute I11_13=rnd(8*B11_13)/8.
compute I12_16=rnd(2*B12_16)/2.
compute RH=rnd(HUMID/20)*20.
compute Fines=rnd(pSize/20)*20.
formats RH (f3.0).
formats Fines (f3.0).
execute.
if RH gt 100 RH = 100.
*string s_events (a25).
*compute s_events = CONCAT(RTRIM(etype),' RH ',rtrim(STRING(RH,f3)),' Fines ','rtrim(STRING(Fines,f3)) ).
execute.
* show probabilities and API percentiles for ALL, Incident and pollutants.
USE ALL.
* FREQUENCIES VARIABLES=etype RH Fines /BARCHART PERCENT /FORMAT=DFREQ
 /ORDER=ANALYSIS.
* TABLE 1
======================================================================.
FREQUENCIES VARIABLES=Hour SO2 to PM25 API HUMID pSize /FORMAT=NOTABLE /PERCENTILES=50 95.0 97.5 99 99.99 100 /ORDER=ANALYSIS.
*===================================================

COMPUTE filter_$=event GT 1.
VARIABLE LABELS filter_$ 'event > 0 (FILTER)'.
VALUE LABELS filter_ $ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.
* TABLE 2
FREQUENCIES VARIABLES=etype /BARCHART PERCENT /FORMAT=DFREQ /ORDER=ANALYSIS.
MEANS TABLES=API by etype /CELLS= COUNT NPCT STDDEV MEAN MAX.
* =======================================================================

* FREQUENCIES VARIABLES=API /FORMAT=NOTABLE /PERCENTILES=50 95.0 97.5 100
 /ORDER=ANALYSIS.
COMPUTE filter_$=(event GT 1 and ( etype EQ 'PM10' or etype EQ 'PM2.5' or etype EQ 'NO' or etype EQ 'NO2' or etype EQ 'O3' or etype EQ 'CO' )).
* or etype EQ 'NOx' or etype EQ 'TSP' or etype EQ 'TSP+NOx+CO'.
FILTER BY filter_ $.
EXECUTE.
* FREQUENCIES VARIABLES=etype /BARCHART PERCENT /FORMAT=DFREQ /ORDER=ANALYSIS.
* FIGURE 2

EXAMINE VARIABLES=API BY etype BY SiteID /PLOT=BOXPLOT /STATISTICS=NONE /NOTOTAL.
EXAMINE VARIABLES=API BY etype BY Fines /PLOT=BOXPLOT /STATISTICS=NONE /NOTOTAL.
EXAMINE VARIABLES=API BY etype BY RH /PLOT=BOXPLOT /STATISTICS=NONE /NOTOTAL.

* Development of a pollution index using IR from Himawari-8
38
Development of a pollution index using IR from Himawari-8

* FIGURE 3

USE ALL.
COMPUTE filter_S=(etype EQ 'PM10' or etype EQ 'PM2.5' or etype EQ 'NO' or etype EQ 'NO2' or etype EQ 'SO2' or etype EQ 'O3')
and (HUMID GT 1 and HUMID LT 99 and pSize GT 1 and pSize LT 99 and API GT 0 and API LT 15).
FILTER BY filter_S.
EXECUTE.
EXAMINE VARIABLES=B7_8 BY etype BY severity /PLOT=BOXPLOT /STATISTICS=DESCRIPTIVES /NOTOTAL.
EXAMINE VARIABLES=B14_15 BY etype BY severity /PLOT=BOXPLOT /STATISTICS=DESCRIPTIVES /NOTOTAL.
EXAMINE VARIABLES=B9_10 BY etype BY severity /PLOT=BOXPLOT /STATISTICS=DESCRIPTIVES /NOTOTAL.
EXAMINE VARIABLES=B11_13 BY etype BY severity /PLOT=BOXPLOT /STATISTICS=DESCRIPTIVES /NOTOTAL.
EXAMINE VARIABLES=B12_16 BY etype BY severity /PLOT=BOXPLOT /STATISTICS=DESCRIPTIVES /NOTOTAL.

* Determine poll ID (priors).
compute poll=0.
formats poll (f3.0).
if (event=8) poll = 0.
if (event=16) poll = 5.
if (event=32) poll = 10.
if (event=64) poll = 20.
if (event=4) poll = 30.
if (event=2) poll = 50.
execute.

* Determine pollution index

* manually looped this multiple times ~20 times.
DELETE VARIABLES pPoll.
execute.
REGRESSION /MISSING LISTWISE /STATISTICS COEFF OUTS R ANOVA CHANGE /CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN /DEPENDENT poll /METHOD=ENTER B14_15 B9_10 B12_16 /save PRED (pPoll).
compute poll=pPoll.
execute.
EXAMINE VARIABLES=pPoll BY etype /PLOT=BOXPLOT /STATISTICS=DESCRIPTIVES /NOTOTAL.
DELETE VARIABLES pPoll.
execute.
compute pPoll  = 17.258 +  4.687 * B14_15 +  0.499 * B9_10 +  0.577 * B12_16.
execute.
EXAMINE VARIABLES=pPoll BY etype by Fines /PLOT=BOXPLOT /STATISTICS=DESCRIPTIVES /NOTOTAL.
EXAMINE VARIABLES=pPoll BY etype by RH /PLOT=BOXPLOT /STATISTICS=DESCRIPTIVES /NOTOTAL.

* show what happens with all ===========================================.
use all.
compute pPoll  = 17.258 +  4.687 * B14_15 +  0.499 * B9_10 +  0.577 * B12_16.
compute iPoll=1*RND(pPoll/1).
formats iPoll (f3.0).
COMPUTE filter_S=(iPoll GT 1) and (iPoll LT 53).
FILTER BY filter_S.
EXECUTE.
EXAMINE VARIABLES=API by iPoll /PLOT=BOXPLOT /STATISTICS=DESCRIPTIVES /NOTOTAL.
EXAMINE VARIABLES=HUMID by iPoll /PLOT=BOXPLOT /STATISTICS=DESCRIPTIVES /NOTOTAL.
EXAMINE VARIABLES=pSize by iPoll /PLOT=BOXPLOT /STATISTICS=DESCRIPTIVES /NOTOTAL.
EXAMINE VARIABLES=NO by iPoll /PLOT=BOXPLOT /STATISTICS=DESCRIPTIVES /NOTOTAL.
EXAMINE VARIABLES=NO2 by iPoll /PLOT=BOXPLOT /STATISTICS=DESCRIPTIVES /NOTOTAL.
EXAMINE VARIABLES=PM25 by iPoll /PLOT=BOXPLOT /STATISTICS=DESCRIPTIVES /NOTOTAL.
EXAMINE VARIABLES=PM10 by iPoll /PLOT=BOXPLOT /STATISTICS=DESCRIPTIVES /NOTOTAL.
EXAMINE VARIABLES=SO2 by iPoll /PLOT=BOXPLOT /STATISTICS=DESCRIPTIVES /NOTOTAL.
EXAMINE VARIABLES=O3 by iPoll /PLOT=BOXPLOT /STATISTICS=DESCRIPTIVES /NOTOTAL.

* FIGURE 6 ========================================================================
* summarize per pollutant.
Use All.
VARIABLE LABELS iPoll 'Pollutant Index'.
DATASET NAME TmpMain WINDOW=FRONT.
COMPUTE filter_$=etype EQ 'NO'.
FILTER BY filter_$.
EXECUTE.
DATASET DECLARE Aggr.
SORT CASES BY iPoll etype.
AGGREGATE /OUTFILE='Aggr' /PRESORTED /BREAK=iPoll etype /Concentration=MEAN(NO) /Counts=N.
DATASET ACTIVATE TmpMain.

Use All.
COMPUTE filter_$=etype EQ 'NO2'.
FILTER BY filter_$.
EXECUTE.
DATASET DECLARE Tmp.
SORT CASES BY iPoll etype.
AGGREGATE /OUTFILE='Tmp' /PRESORTED /BREAK=iPoll etype /Concentration=MEAN(NO2) /Counts=N.
DATASET ACTIVATE Aggr.
ADD FILES /FILE=* /FILE='Tmp'.
EXECUTE.
DATASET ACTIVATE TmpMain.
DATASET CLOSE Tmp.

Use All.
COMPUTE filter_$=etype EQ 'PM2.5'.
FILTER BY filter_$.
EXECUTE.
DATASET DECLARE Tmp.
SORT CASES BY iPoll etype.
AGGREGATE /OUTFILE='Tmp' /PRESORTED /BREAK=iPoll etype /Concentration=MEAN(PM25) /Counts=N.
DATASET ACTIVATE Aggr.
ADD FILES /FILE=* /FILE='Tmp'.
EXECUTE.
DATASET ACTIVATE TmpMain.
DATASET CLOSE Tmp.

Use All.
COMPUTE filter_$=etype EQ 'PM10'.
FILTER BY filter_$.
EXECUTE.
DATASET DECLARE Tmp.
SORT CASES BY iPoll etype.
AGGREGATE /OUTFILE='Tmp' /PRESORTED /BREAK=iPoll etype /Concentration=MEAN(PM10) /Counts=N.
DATASET ACTIVATE Aggr.
ADD FILES /FILE=* /FILE='Tmp'.
EXECUTE.
DATASET ACTIVATE TmpMain.
DATASET CLOSE Tmp.
Use All.
COMPUTE filter_\_$=etype EQ 'SO2'.
FILTER BY filter_\_.$.
EXECUTE.
DATASET DECLARE Tmp.
SORT CASES BY iPoll etype.
AGGREGATE /OUTFILE='Tmp' /PRESENTED /BREAK=iPoll etype /Concentration=MEAN(SO2) /Counts=N.
DATASET ACTIVATE Aggr.
ADD FILES /FILE=* /FILE='Tmp'.
EXECUTE.
DATASET ACTIVATE TmpMain.
DATASET CLOSE Tmp.

Use All.
COMPUTE filter_\_$=etype EQ 'O3'.
FILTER BY filter_\_.$.
EXECUTE.
DATASET DECLARE Tmp.
SORT CASES BY iPoll etype.
AGGREGATE /OUTFILE='Tmp' /PRESENTED /BREAK=iPoll etype /Concentration=MEAN(O3) /Counts=N.
DATASET ACTIVATE Aggr.
ADD FILES /FILE=* /FILE='Tmp'.
EXECUTE.
DATASET ACTIVATE TmpMain.
DATASET CLOSE Tmp.

DATASET ACTIVATE Aggr.
SORT CASES BY iPoll.
AGGREGATE /OUTFILE=* MODE=ADDVARIABLES /PRESENTED /BREAK=iPoll
/Counts_sum=SUM(Counts).
compute Probability=Counts/Counts_sum.
execute.

GRAPH /SCATTERPLOT(BIVAR)=iPoll WITH Concentration BY etype /MISSING=LISTWISE.
GRAPH /SCATTERPLOT(BIVAR)=iPoll WITH Probability BY etype /MISSING=LISTWIS.
DATASET ACTIVATE TmpMain.
DATASET CLOSE Aggr.

* now setup dummy events for RH, Fines, Low, Med, High
* Give a text label to the classification.
use all.
string Tag(a10).
compute Tag =STRING(event,N3).
compute Tag="RH".

COMPUTE filter_\_$=Tag EQ 'RH' and API > 0.
FILTER BY filter_\_.$.
EXECUTE.
DATASET DECLARE Aggr.
SORT CASES BY iPoll Tag.
AGGREGATE /OUTFILE='Aggr' /PRESENTED /BREAK=iPoll Tag /Concentration=MEAN(RH) /Counts=N.
DATASET ACTIVATE TmpMain.
compute Tag="Fines".

Use All.
COMPUTE filter_\_$=Tag EQ 'Fines' and API > 0.
FILTER BY filter_\_.$.
EXECUTE.
DATASET DECLARE Tmp.
SORT CASES BY iPoll Tag.
AGGREGATE /OUTFILE='Tmp' /PRESENTED /BREAK=iPoll Tag /Concentration=MEAN(pSize) /Counts=N.
DATASET ACTIVATE Aggr.
ADD FILES /FILE=* /FILE='Tmp'.
EXECUTE.
DATASET ACTIVATE TmpMain.
DATASET CLOSE Tmp.

use all.
if (API > 0 and API < 1.416) TAG = 'Low'.
if (API > 1.416 and API < 3.78) TAG = 'Med'.
if (API > 3.78 ) TAG = 'High'.

Use All.
COMPUTE filter_$=Tag EQ 'Low'.
FILTER BY filter_.
EXECUTE.
DATASET DECLARE Tmp.
SORT CASES BY iPoll Tag.
AGGREGATE /OUTFILE='Tmp' /PRESORTED /BREAK=iPoll Tag /Concentration=MEAN(API) /Counts=N.
DATASET ACTIVATE Aggr.
ADD FILES /FILE=* /FILE='Tmp'.
EXECUTE.
DATASET ACTIVATE TmpMain.
DATASET CLOSE Tmp.

Use All.
COMPUTE filter_$=Tag EQ 'Med'.
FILTER BY filter_.
EXECUTE.
DATASET DECLARE Tmp.
SORT CASES BY iPoll Tag.
AGGREGATE /OUTFILE='Tmp' /PRESORTED /BREAK=iPoll Tag /Concentration=MEAN(API) /Counts=N.
DATASET ACTIVATE Aggr.
ADD FILES /FILE=* /FILE='Tmp'.
EXECUTE.
DATASET ACTIVATE TmpMain.
DATASET CLOSE Tmp.

Use All.
COMPUTE filter_$=Tag EQ 'High'.
FILTER BY filter_.
EXECUTE.
DATASET DECLARE Tmp.
SORT CASES BY iPoll Tag.
AGGREGATE /OUTFILE='Tmp' /PRESORTED /BREAK=iPoll Tag /Concentration=MEAN(API) /Counts=N.
DATASET ACTIVATE Aggr.
ADD FILES /FILE=* /FILE='Tmp'.
EXECUTE.
DATASET ACTIVATE TmpMain.
DATASET CLOSE Tmp.

DATASET ACTIVATE Aggr.
SORT CASES BY iPoll.
AGGREGATE /OUTFILE=* MODE=ADDVARIABLES /PRESORTED /BREAK=iPoll
/Counts_sum=SUM(Counts).
compute Probability=3*Counts/Counts_sum.
execute.
GRAPH /SCATTERPLOT(BIVAR)=iPoll WITH Concentration BY Tag /MISSING=LISTWISE.
GRAPH /SCATTERPLOT(BIVAR)=iPoll WITH Probability BY Tag /MISSING=LISTWISE.
DATASET CLOSE Aggr.