Authors:

Group 18

Thomas Koh 329888

Vishall Krishnan 1018473

Shaohua Liu 1150336

Yujing Yang XXX

Abstract

The planetary boundary layer (PBL) height is an important factor for measuring the pollution level of a given area. The DPIE have asked us to conduct an analysis comparing the PBL height values based on two different sources, the CL51 ceilometer and the CT model output. This report shows our preliminary exploratory data analysis conducted on the data and our proposal going forward

Comparison of pbl height based on cl51 ceilometer and ct model outputs in merriwa and lidcombe

An EDA analysis and scoped proposal

Contents

[A. Introduction 2](#_Toc72664631)

[Background: 2](#_Toc72664632)

[Problem statement: 2](#_Toc72664633)

[Intended outcomes of analysis: 2](#_Toc72664634)

[B. Literature Review and Motivation 3](#_Toc72664635)

[History and development of air quality models 3](#_Toc72664636)

[The importance of modelling pollution levels 3](#_Toc72664637)

[The importance of PBL in pollution modelling 4](#_Toc72664638)

[On the Vaisala CL51 ceilometer 4](#_Toc72664639)

[C. Data analysis and preliminary modelling 5](#_Toc72664640)

[Data provided 5](#_Toc72664641)

[Data pre-processing steps 5](#_Toc72664642)

[D. Analysis 1 - Comparing the PBL height of both datasets. 7](#_Toc72664643)

[Visualisation of the CL51 and CTM PBL height data 7](#_Toc72664644)

[Statistical metrics 9](#_Toc72664645)

[Discussion 10](#_Toc72664646)

[E. Analysis 2 - Analysing factors that impact the PBL heights 11](#_Toc72664647)

[Statistical metrics 11](#_Toc72664648)

[Visualization of PBL height with the external factors 11](#_Toc72664649)

[Comparing PBL heights against external factors in both locations 11](#_Toc72664650)

[Comparing PBL heights against external factors in Merriwa 12](#_Toc72664651)

[Comparing PBL heights against external factors in Lidcombe 13](#_Toc72664652)

[F. Proposal for Semester 2 15](#_Toc72664653)

[G. Proposed timeline for Semester 2 15](#_Toc72664654)

[H. References 16](#_Toc72664655)

[Appendix A: Details on dataset provided. 17](#_Toc72664656)

[Appendix B: Formulae for statistical metrics 18](#_Toc72664657)

[Appendix C: Description of external factors considered 20](#_Toc72664658)

[Appendix D: Correlation tables between PBL heights and external factors 21](#_Toc72664659)

## Introduction

### Background:

There has traditionally been a lot of pollution in New South Wales (NSW) due to mining in particular from mining dust. As such, there has been concerted effort to better monitor these pollution sites. As part of this effort The Department of Planning, Industry and Environment of the NSW Government (DPIE) have obtained new instruments as part of an industry programme to monitor pollution levels.

One such instrument is the CL51 ceilometer by Vaisala which is an air quality measurement tool (Vaisala, 2020). This is rolled out in two locations: (1) Merriwa and (2) Lidcombe. We have therefore been asked to investigate the efficacy of a new air quality measuring tool in measuring pollution levels in these two areas.

### Problem statement:

We have been notified by DPIE that the Planetary Boundary Layer Height (PBL height) is an important factor for measuring the pollution level of a given area.

We have subsequently been tasked to compare the quality of the PBL height as measured by the new CL51 ceilometer instrument. In doing so we aim to answer the following questions:

* How accurate and reliable is the PBL height data as obtained from the CL51 ceilometer;
* How do the readings from the CL51 ceilometer compare to the predicted PBL height as obtained from the CT model; and
* What are the potential improvements that can be made to the prediction of PBL height based on information obtained from the CL51 ceilometer.

For each of the two locations, we are presented with the following information:

* CL51 data for the period from 12/2/2021 to 19/2/2021; and
* Predicted PBL height values;

### Intended outcomes of analysis:

In Semester 1, our focus was on performing detailed exploratory data analysis on the PBL height data provided. The intended outcome of the analysis is as follows:

* To develop metrics for comparing the PBL height from the CL51 ceilometer against the predicted PBL height data from the CT model output; and
* To understand the factors that drive the change in PBL height.

## Literature Review and Motivation

### History and development of air quality models

Air quality has become one of the most important indicators evaluating the environment since November 1979 when *Convention on long-range Transboundary Air Pollution* was concluded.

Air quality models have undergone various permutations over the course of the last 60 years. Xue et al. (2013) has categorised three broad model classes of air quality models in place during the 1970s to 1980s in the US – so called first-generation models. They are: 1) Box models - which are based on conservation of mass, 2) Gauss diffusion model, which is based on statistical theory of turbulent diffusion - Industry Source Complex (ISC), AMS/EPA Regulatory Model (AERMOD) and Atmospheric Dispersion Modeling System (ADMS) are all classic Gauss diffusion models and 3) Joseph-Louis Lagrange model which are represented models are OZIP/EKMA and CALPUFF.

These first-generation models were then further improved upon in the 1980s by incorporating urban and regional environmental factors, such as particles and photochemical smog, which are found to have close relations with air quality. These gave rise to second-generation Euler Grid models which includes models such as the Urban airshed model (UAM), ROM and Regional Acid Deposition Model (RADM).

Further improvements to the second-generation models were made in the late 1990s to incorporate interactions between regional factors. These led to the development of third-generation models such as the CMAQ (Community Multiscale Air Quality model developed by the US EPA). The CMAQ is “a three-dimensional Euler model with multi-module integration and multi-scale nesting” (Xue et al., 2013, p. 15). Crucial to the implementation of these third-generation models is the concept of “one atmosphere”, which attempts to model the atmosphere as one cohesive object.

Xue et al. (2013) also mentioned that apart from the CMAQ model, CAMX (Comprehensive Air quality Model with extensions developed by ENVIRON)and WRF-Chem (Weather Research and Forecasting   model coupled with Chemistry) are widely used in monitoring air quality nowadays.

WRF-CHEM, developed by the American National Center for Atmospheric Research, is considered the state of the art. This is because its simulations come close to the real atmosphere, such that the chemical and physical processes are coupled (Boxe et al. 2016). This results in “… air quality component of the model [that] is fully consistent with the meteorological component; both components use the same transport scheme (mass and scalar preserving), the same grid (horizontal and vertical components), and the same physics schemes for subgrid-scale transport. The components also use the same timestep, hence no temporal interpolation is needed” (Grell et al., 2005). It is the first fully compressible conservative-form nonhydrostatic atmospheric model suitable for both research and weather prediction applications.

### The importance of modelling pollution levels

After the 68th General Assembly of the World Health Organization adopted a resolution on the health effects of air pollution, a great many governments devoted to creating the air quality model related to their special geographical environment, including the Australian Government.

Duc et al. (2015) in particular noted that New South Wales (NSW) has a unique geographical environment. This can result in particles such as dust, sea salt, and wildfire aerosols being transported across wide regions of the state. This gave rise to the need for an integrated CTM framework that includes models that are flexible and that can be scaled to different resolutions.

The CCAM (Cubic Conformal Atmospheric Model) is one such model that allows dynamical downscaling of global meteorological forecast or analysis data and is therefore suitable for the need of flexibility in NSW’s model. The CCAM “… achieves high efficiency as a result of using semi-Lagrangian, semi-implicit time differencing. It is formulated on a quasi-uniform grid, derived by projecting the panels of a cube onto the surface of the Earth” (McGregor & Dix, 2008, p. 51).

In practice, the CCAM is incorporated into the CTM framework to produce “… downscaled meteorology data for four Australian grid domains at 80kmx80km, 27kmx27km, 9kmx9km and 3kmx3km resolution for use in CTM simulations” (Duc et al., 2015, CCAM-CTM modeling system).

### The importance of PBL in pollution modelling

The basic idea of the CCAM-CTM is to determine the planetary boundary layer (PBL) height. The PBL height “… mainly determines the environmental capacity for the diffusion of atmospheric pollutants” (Shi et al. 2020, p. 1), which is to say is an important marker of air pollution. As such, it comes as no surprise that the PBL height (also known as mixing height) forms a key factor in air pollution models. Despite this, there are no direct methods to determine the PBL height (Eresmaa et al. 2006, p. 1485).

Eresmaa et al. (2006) further noted that the ceilometer is one such instrument that can be used to estimate the PBL height. It uses the lidar technique and measures the aerosol concentration profile of the atmosphere. The ceilometer reports the cloud base height, vertical visibility and cloud cover at a given location and can operate unattended in harsh weather conditions (Münkel & Roininen, 2008, introduction).

### On the Vaisala CL51 ceilometer

Vaisala is a company that specialise in the manufacturing of environmental measurement devices. The Vaisala ceilometers are widely used and have been evolving these years - from the CT12K model to the more recent CL51 model.

The readings obtained from the CL51 ceilometer can be used to provide information on the PBL and is also used for atmospheric analysis. In particular, the CL51 uses the pulsed diode laser LIDAR (light detection and ranging) technology which enables the device to detect up to 3 cloud layers simultaneously. It can also provide a backscatter profile of the measurement range (Vaisala, 2020, measurement from ground level).

According to Munkel and Roininen (2008), the backscatter profile as obtained from the ceilometer readings can be used to track aerosol concentrations, which in turn can be used to determine the structure of the PBL heights. One such method to do this is the gradient method which “… looks for the steepest decrease within the backscatter profile. In most cases the lowest of gradient minima marks the top of the mixed layer” (Münkel & Roininen, 2008, method).

## Data analysis and preliminary modelling

In this section we outline the results of our preliminary analysis of the PBL height data provided.

### Data provided

DPIE has provided the following PBL height data for the period between 12/2/2021 to 19/2/2021 at Merriwa and Lidcombe:

* Data from the Vaisala CL51 ceilometer; and
* Data from the CT model.

Both datasets are provided in the form of CSV files.

A detailed description of the dataset provided is set out in [Appendix A](#_Appendix_A:_Details).

### Data pre-processing steps

We have used Python to pre-process the data. The following sets out the detailed steps taken to pre-process both datasets provided.

#### Pre-processing the CL51 dataset

1. Data is firstly loaded into Python.
2. We notice that missing values in the CSV files are automatically filled with “-999”. This is changed into ‘NaN’ values for ease of future processing.
3. A proper datetime stamp is created based on information already present in the data.
4. Imputation is performed on the “bl\_height” feature to append the missing values. We have chosen to impute the missing values (as opposed to omitting the data with missing values) because we have limited data to work with (i.e., only one week’s worth of observations). We have used the Pandas library’s built-in interpolation method (interpolated based on the datetime stamp).

#### Pre-processing the CT model data

1. Data is again similarly loaded into Python.
2. We note that the “time” field recorded in this dataset is in UTC. We have adjusted this datafield to align the timestamp with the CL51 data which is in AEST. This is done by adding 10 hours to the data in the “time” field.
3. A datetime stamp is then created using the revised “time” field as described above.

#### Aligning the two datasets

We further note that the granularity of the CL51 dataset datetime stamp is in seconds, whereas the granularity of the CT model output datetime stamp is in hours.

To make the two datasets comparable, we have grouped the CL51 dataset such that it has the same datetime stamp granularity as the CT model outputs (i.e. hours as opposed to minutes). As such in the grouped CL51 dataset, the PBL height at each hour reflects the average PBL height within that hour from the original dataset.

This is then combined with the pre-processed CT model data and used for comparisons.

#### Comparing the CL51 datasets before and after pre-processing

The following graphs show the PBL height as obtained from the CL51 ceilometer in Merriwa and Lidcombe before and after the datasets were imputed.

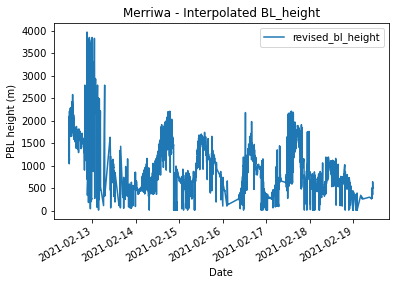
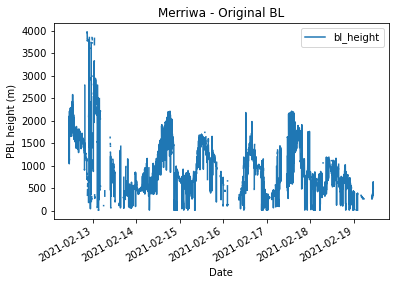


Figure Original vs interpolated PBL heights in Merriwa

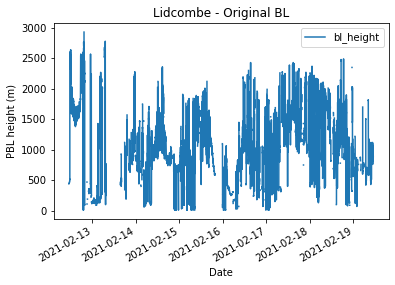
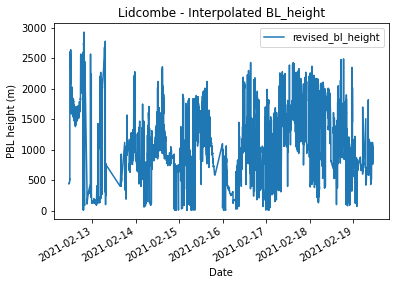


Figure Original vs interpolated PBL heights in Lidcombe

We observe that the pre-processing has successfully filled in the gaps in the missing values in both locations. We further note that based on visual inspection, the method of interpolating the data between datetime stamps is an appropriate in these circumstances given the granularity of the data originally provided.

## Analysis 1 - Comparing the PBL height of both datasets.

In this section we consider a detailed exploratory data analysis of the PBL height as presented in both the CL51 and CT model output datasets.

This section is divided into two parts:

1. Visual comparisons of the CL51 and CTM PBL height data; and

2. Derivation of statistical metrics for comparison of both datasets.

### Visualisation of the CL51 and CTM PBL height data

We firstly denote the CL51 PBL height as the actual PBL height observed in an area. Consequently, the CT modelled PBL height is then denoted as the predicted PBL height.

The following graphs show a comparison between the CL51 (actual) PBL height and the CT model (predicted) PBL height at both locations.

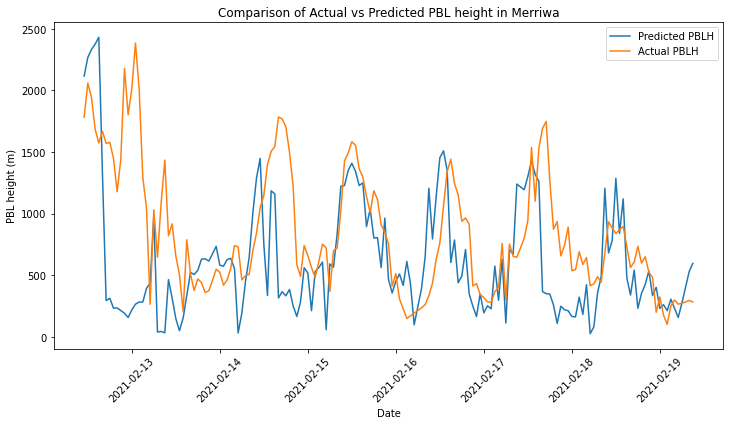


Figure 3 Actual vs Predicted PBL height in Merriwa

Overall, in Merriwa we observe that there is quite a tight overlap between predicted and actual PBL height over the one-week period.

We note however observations in the timeframe from 12/2/2021 to 13/2/2021 deviate significantly. This is due to poor quality readings from CL51 ceilometer which may be as a result of rain during this period.

From midday 13/2/2021 to around midday 17/2/2021 the CT model outputs was in agreement with the actual observed PBL height. There is then another discrepancy in readings from midday 17/2/2021 to 18/2/2021. Unsure as to why this is the case

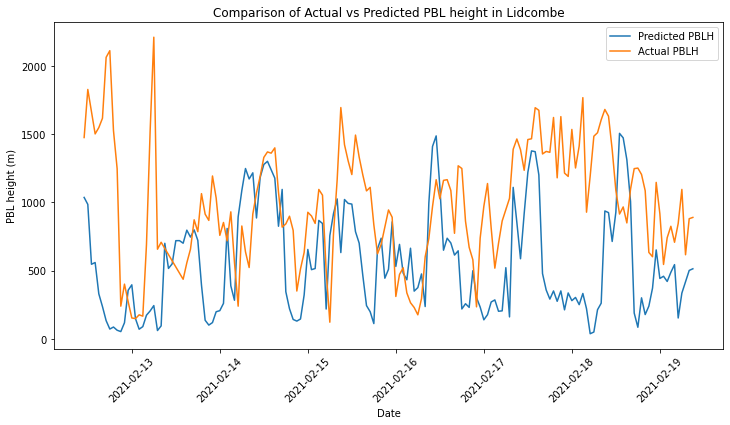


Figure 4 Actual vs Predicted PBL height in Lidcombe

In Lidcombe however, the CT model outputs and CL51 ceilometer readings exhibit significantly more deviations.

We observe that while there is some agreement as to the timing of the peaks and troughs of the PBL height as measured by the CT model and the CL51 ceilometer, their magnitudes are vastly different. For instance, the actual versus predicted PBL heights at 13/2/2021 and again from 17/2/2021 to 18/2/2021 are in total opposition to one another.

The following graphs compares the PBL height at each hour of the day grouped by the date of the observations.

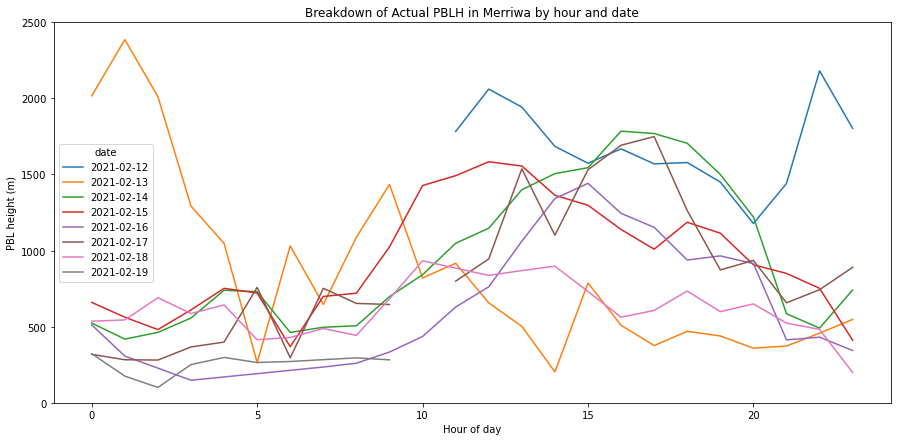


Figure 5 Breakdown of actual PBL height in Merriwa by hour

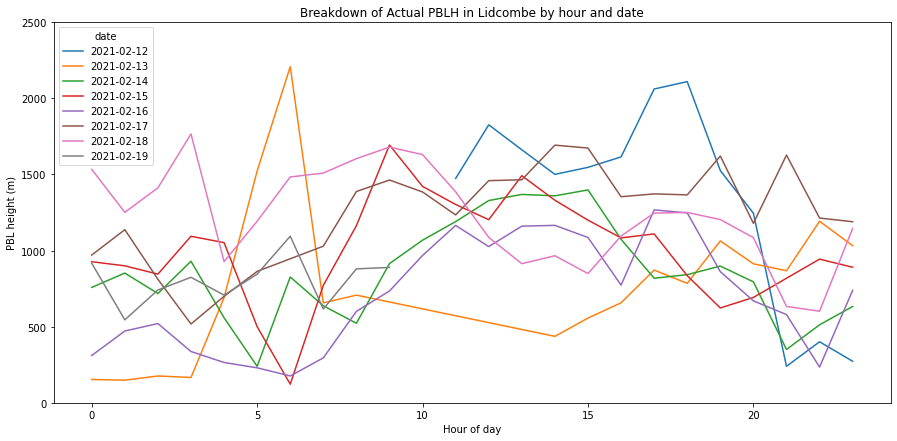


Figure 6 Breakdown of actual PBL height in Lidcombe by hour

The above graphs further highlight the cyclic nature of the PBL height. We observe that the PBL height tends to reach peak height during the middle of the day (at around 3pm) before dropping off again and troughing at around 5 am.

### Statistical metrics

In addition to visual inspections, we have also calculated some select statistical metrics aimed at comparing the fit between the actual PBL height and the predicted PBL height. To do this, we have selected 5 metrics which are as follows:

* Wilmot’s index of agreement;
* Root mean squared error (RMSE);
* Mean absolute error;
* Mean bias error; and
* DTW distance

A detailed derivation of the statistical measures used and their interpretations are set out In Appendix B.

The following sets out a summary of the statistical metrics measured for both locations:

|  |  |  |
| --- | --- | --- |
|  | **Location** | |
|  | **Merriwa** | **Lidcombe** |
| Wilmot's Index of agreement | 0.64061 | 0.47060 |
| RMSE | 597 | 684 |
| Mean absolute error | 416 | 540 |
| Mean bias error | -233 | -430 |
| DTW distance | 91,526 | 107,124 |

The measures indicate that there is a closer fit between the actual and predicted PBL heights in Merriwa compared to Lidcombe.

### Discussion

The PBL height exhibits a peak and trough type behaviour where peaks are typically around midday and troughs are late at night/early mornings. Overall, there is a fairly reasonable correlation between the predicted and actual PBL over the one-week period. The CT model was able to capture the behaviour of the PBL height in both locations.

There are however significant deviations in the quality of the predicted PBL height from the CT model as compared to the actual PBL heights from the CL51 ceilometer in both locations. In particular, we note that in Merriwa the CT model outputs tend to match the actual PBL height much more closely than in Lidcombe.

While we offer no concrete reasons as to why this is the case, we reason that this may be driven in large part by the urban and geographic environment of the two sets of observations. Lidcombe, being an inner suburb within the Sydney Metropolitan area, is subject to constant pollution and disturbances from urban infrastructure that may potentially impact the reading of the ceilometer. Merriwa by contrast, is located in the Hunter Valley where there are little disturbances from urban infrastructures.

## Analysis 2 - Analysing factors that impact the PBL heights

In this section we analyse external factors that potentially influence the actual and predicted PBL heights in both locations. The set of external factors considered in this analysis is set out in Appendix C.

### Statistical metrics

When comparing the PBL heights against the external factors, we have used two statistical measures:

* Pearson’s correlation; and
* P-value.

We use the above measurements to check each external factor’s correlation with predicted and actual PBL height and p-value. A detailed description of the formulae is set out in Appendix B.

The correlation tables for each location against the external factors are set out in Appendix D.

### Visualization of PBL height with the external factors

In addition to the statistical metrics, plots were also done to validate the correlation calculated between the boundary layer heights and the various external factors. The plots are done in two-fold by comparing the PBL heights with:

1. Factors relating to weather such as rainfall, temperature etc.
2. Factors relating to the chemical particles present in the atmosphere such as carbon monoxide, sulphur dioxide etc.

The plots of the PBL heights and the external factors are used to validate the statistical inferences inferred below.

### Comparing PBL heights against external factors in both locations

We observe that external factors of humidity, ozone and temperature have an influence on the PBL height in both locations.

Specifically:

* Humidity has a negative correlation with PBL height;
* Ozone and temperature have positive correlations with PBL height.

Besides, nitrogen dioxide and wind speed (10m) are not correlated with the actual PBL height of CL51 (as measured by their p-values), but these external factors are correlated with predicted PBL height of CT model.

Moreover, PM10 and wind direction (10m) show positive relationships with actual PBL height, but they are not correlated with predicted PBL height.

Last but not the least, wind direction sigma theta shows contrary relation. It has a weak positive relationship with actual PBL height but weak negative relationship with predicated PBL height.

### Comparing PBL heights against external factors in Merriwa

We observe that humidity remains strongly negatively correlated with PBL heights. Meanwhile, ozone and temperature remain strongly positively correlated with PBL heights. However, there is contrary to the overall analysis.

Specifically:

* Carbon monoxide is positively correlated to PBL heights in Merriwa;
* Wind direction sigma Theta has no relationship with PBL heights in Merriwa.

Since solar radiation information is not available for Merriwa, no conclusion can be drawn with respect to this external factor.

The following plots show the plots of comparing the PBL heights with the external factors in Merriwa.

* A picture containing chart

  Description automatically generatedA picture containing chart

  Description automatically generatedHumidity is shown to be strongly negatively correlated with the PBL heights.

Figure 7 Actual and Predicted PBL vs humidity in Merriwa

* Ozone is shown to be strongly positively correlated with the BL heights.

A picture containing text

Description automatically generatedA picture containing whiteboard

Description automatically generated

Figure 8 Actual and Predicted PBL vs Ozone in Merriwa

* Temperature is shown to be positively correlated with the PBL heights.

A picture containing chart

Description automatically generatedA picture containing chart

Description automatically generated

Figure 9 Actual and Predicted PBL vs temperature Merriwa

### Comparing PBL heights against external factors in Lidcombe

The statistical metric of Lidcombe indicates humidity remains negatively correlated with PBL heights in Lidcombe. In addition, solar radiation and wind speed (10m) remain strong positive correlations with predicted PBL heights. Temperature remains positively correlated with actual PBL heights. Furthermore, Nitrogen Dioxide is not significantly correlated with the actual PBL height (as measured by their p-values) but is however significantly correlated with predicted PBL height of the CT model. By contrast with overall analysis, PM10 is not correlated with actual PBL height in Lidcombe.

The following plots show the plots of comparing the PBL heights with the external factors in Lidcombe.

* Humidity is shown to be negatively correlated with the PBL heights.

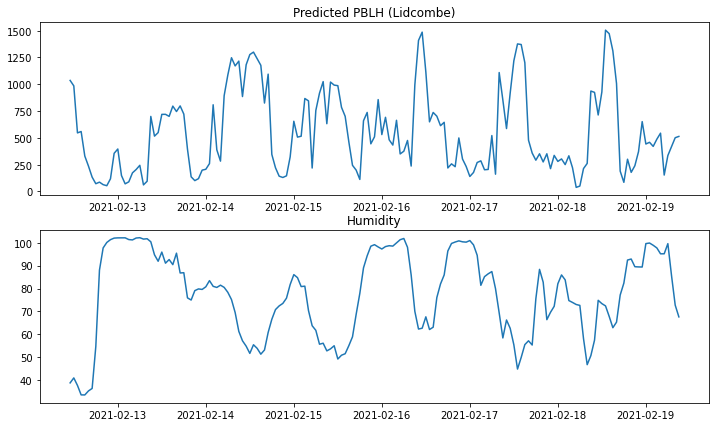
CL51


Figure 10 Actual and Predicted PBL vs humidity in Lidcombe

* Solar radiation is shown to be positively correlated with the PBL heights.

A picture containing text

Description automatically generatedA picture containing text

Description automatically generated

Figure 11 Actual and Predicted PBL vs solar radiation in Lidcombe

* Temperature is shown to be positively correlated with the PBL heights.

Text

Description automatically generated with medium confidenceText

Description automatically generated

Figure 12 Actual and Predicted PBL vs temperature in Lidcombe

* Nitrogen dioxide is negatively correlated with the predicted PBL height but is not very correlated with the actual BL height.

Text

Description automatically generatedA picture containing text

Description automatically generated

Figure 13 Actual and Predicted PBL vs nitrogen dioxide in Lidcombe

## Proposal for Semester 2

Semester 2 will focus on the continuation of Analysis 2. In particular:

1. Further in-depth analysis of the factors that drive PBL height;
2. Derivation of model(s) that may be used to predict the PBL height based on such factors; and
3. Analysis of our derived model against: (1) the original PBL height data provided and (2) CT model output provided.

We also aim to provide a report summarizing our findings for DPIE.

## Proposed timeline for Semester 2

|  |  |  |
| --- | --- | --- |
| Task | Timeline in Semester 2 | Doers |
| Further in-depth analysis of factors that drive the PBL height | Weeks 1 – 2 | Vishall and Shaohua |
| Derivation of models to predict the PBL height based on these factors | Weeks 1 – 4 | All members |
| Analysis of models developed against CL51 data | Weeks 3 – 5 | Shaohua and Yujing |
| Analysis of models developed against the CT model output | Weeks 3 – 5 | Thomas and Vishall |
| Consultations with DPIE in relation to our findings | Week 6 | All members |
| Further revisions to the modelling based on feedback from DPIE | Weeks 6 – 9 | All members |
| Final write-up of report and findings | Weeks 9 – 12 | All members |
| Further consultations with DPIE as required | Weeks 10 – 11 | All members |
| Final presentation of reports and findings to DPIE and cohort | Weeks 11 – 12 | All members |

## References

[1] Boxe, C., Hafsa, U., Blue, S., Emmanuel, S., Griffith, E., & Moore, J., et al., (2016). *Air Quality Modeling and Forecasting over the United States Using WRF-Chem*. AGU Fall Meeting, AGU Fall Meeting Abstracts.

[2] CL51 Ceilometer for High-Range Cloud Height Detection, (2021).[online] Available at <<https://www.vaisala.com/sites/default/files/documents/CL51-Datasheet-B210861EN.pdf>> (23/05/2021 02:36)

[3] Duc, H., Trieu, T., Scorgie, Y., Cope, M., & Thatcher, M., (2015). *Air Quality Modelling of The Sydney Region Using CCAM-CTM.* Clean Air Society of Australia & New Zealand (CASANZ) 22th International Clean Air and Environment Conference.

[4] Eresmaa, N., Karppinen, A., Joffre, S. M., Räsänen, J., & Talvitie, H. (2006). *Mixing height determination by ceilometer*. Atmospheric Chemistry and Physics, 6(6), 1485-1493.

[5] Grell, G.A., Peckham, S.E., Schmitz, R., McKeen, S.A., Frost, G., Skamarock, W.C., & Eder, B., (2005). *Fully coupled “online” chemistry within the WRF model*. Atmospheric Environment,37,6957-6975.

[6] McGregor, J.L., & Dix, M.R., (2008). *High Resolution Numerical Modelling of the Atmosphere and Ocean.* New York: Springer

[7] Münkel, C. & Roininen, R. (2010). *Investigation of boundary layer structures with ceilometer using a novel robust algorithm*. American Meteorological Society.

[8] Münkel, C., & Roininen, R. (2008, January). *Mixing layer height assessment with a compact lidar ceilometer*. In The 88th Annual Meeting Symposium on Recent Developments in Atmospheric Applications of Radar and Lidar (pp. 20-24).

[9] Shi, Y., Hu, F., Xiao, Z., Fan, G., & Zhang, Z. (2020). *Comparison of four different types of planetary boundary layer heights during a haze episode in Beijing.* Science of the total environment, 711, 134928.

[10] Vaisala, 2020. *Ceilometer CL51 for high-range cloud height detection,* s.l.: Viasala Oyj.

[11] Xue, W., Wang, J., Yang, J., Lei, Y., Wang, Y. & Chen, X., (2013). *Domestic and foreign research* *progress of air quality model*. Environment and Sustainable Development,38,14-20.

## Appendix A: Details on dataset provided.

In respect of the CL51 dataset:

|  |  |
| --- | --- |
| **Data fields** | **Description** |
| # Time | Datetime stamp of the observation in AEST |
| Layer\_QualityIndex | Quality of information read |
| bl\_height | PBL height in metres |
| Mean\_Layer\_Height | Mean of the PBL height in metres |
| n\_BL | Number of PBL observed |
| cloud\_stat | Number of cloud layers observed |
| cloud\_dat | Height of cloud data in metres |

In respect of the CT model dataset:

|  |  |
| --- | --- |
| **Data fields** | **Description** |
| lat | Latitude of observed location |
| lon | Longitude of observed location |
| date | Date of observation |
| time | Time of observation in UTC |
| temperature | Predicted temperature in Celsius |
| mixing\_height | Predicted PBL height in metres |

## Appendix B: Formulae for statistical metrics

As a matter of convenience, we will adopt the following notations:

|  |  |
| --- | --- |
| Notation | Description |
|  | Actual PBL height observed at a particular location at time *i.*  This corresponds to dataset obtained from the CL51 ceilometer. |
|  | Predicted PBL height observed at a particular location at time *i.*  This corresponds to dataset obtained from the CT model output. |
|  | The average actual PBL height observed at a particular location based on the CL51 ceilometer data. |
|  | The average predicted PBL height observed at a particular location based on the CT model output. |
|  | Number of observations |
|  | The value of the external factor observed at a particular location at time *i.* |
|  | The average value of the external factor observed at a particular location at time *i.* |

The following statistical measures have been used in Sections D and E.

|  |  |
| --- | --- |
| **Statistical measures** | **Description of measure** |
| Wilmott’s Index of Agreement | A numeric measure that measures how well a model simulates observed data. Simulations that fit the observed data well will exhibit higher index of agreement values.  The formula is given as follows: |
| Root mean squared error (RMSE) | Measures the root of the average squared difference between the actual and predicted PBL height at a given area. This is a proxy measure for the distance between the actual and predicted PBL heights. This metric naturally gives precedence to larger deviations between the actual and predicted PBL heights.  The formula is given as follows: |
| Mean absolute error | Measures the average absolute difference between the actual and predicted PBL height at a given area. This is another proxy measure for the distance between the actual and predicted PBL heights. Contrary to RMSE, this measure is not skewed by larger deviations between the actual and predicted PBL heights.  The formula is given as follows: |
| Mean bias error | Measures the average difference between the actual and predicted PBL height at a given area.  The formula is given as follows: |
| Dynamic Time Warping (DTW) distance | DTW presents an alternative metric for measuring the distance between the actual and predicted PBL heights.  In contrast to the RMSE, DTW attempts to find the best mapping between the actual and predicted PBL heights such that the distance between the mappings is minimised. Such techniques are called time-warping. |
| Pearson correlation | Measures actual and predicted PBL height whether has a linear relationship with external factors.  For external factors and CL51 (actual) PBL Height:    For external factors and CT Model(predicted) PBL Height: |
| p-value | Measures the correlation between PBL height and external factors is significant or not.  The null hypothesis is coefficient of external factor is 0, which indicates the external factor has no relationship with actual or predicted PBL height.  If p-value less than 0.05, then we reject the null hypothesis, and the correlation is statistically significant.  If the p-value is greater than 0.05, then we cannot conclude that the correlation is different from 0. |

## Appendix C: Description of external factors considered

The following table sets out the set of external factors that are analysed and a brief description of what they are:

|  |  |
| --- | --- |
| External factors | Description[[1]](#footnote-1) |
| Carbon monoxide | Concentration of carbon monoxide in atmosphere reported in parts per million (ppm) |
| Humidity | Relative humidity in atmosphere measured as percentage of maximum humidity. |
| Nephelometer | Measure of visibility by measuring light scattering reported in 10-4 m-1 |
| Nitric Oxide | A principal form of oxides of Nitrogen that is commonly found in the atmosphere measured by pars per hundred million (pphm) |
| Nitrogen Dioxide | A form of oxides of Nitrogen that is also commonly found in the atmosphere measured by pphm |
| Ozone | A gaseous secondary pollutant that is found in the atmosphere measured by pphm |
| PM10 | Fine particles of less than 10 micrometers (μm) in diameter found in the atmosphere measured by μg/m3 |
| PM2.5 | Fine particles of less than 2.5 μm in diameter found in the atmosphere measured by μg/m3 |
| Rainfall | Amount of rainfall measured by mm/m2 |
| Sulphur Dioxide | A form of respiratory pollutant found in the atmosphere measured by pphm |
| Temperature | Temperature measured by Celsius |
| Wind Direction (10m) | Direction that the wind is blowing from measured in degrees |
| Wind Direction Sigma Theta | Standard deviation in the wind direction measured in degrees |
| Wind Speed (10m) | Wind speed measured in m/s |
| Solar radiation | Solar radiation of a given location measured by watts/m2 |

## Appendix D: Correlation tables between PBL heights and external factors

Actual and predicted PBL height in both locations combined vs external factors:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | correlation | | p-value | |
|  | CL51 | CT Model | CL51 | CT Model |
| **Carbon monoxide** | 0.0927 | -0.1846 | 0.1251 | 0.0021 |
| **Humidity** | -0.6672 | -0.5184 | 0 | 0 |
| **Nephelometer** | 0.0069 | -0.1734 | 0.9091 | 0.0039 |
| **Nitric Oxide** | 0.1445 | -0.1256 | 0.0165 | 0.0374 |
| **Nitrogen Dioxide** | 0.0113 | -0.4249 | 0.8522 | 0 |
| **Ozone** | 0.4682 | 0.4554 | 0 | 0 |
| **PM10** | 0.3528 | -0.0240 | 0 | 0.6923 |
| **PM2.5** | 0.0175 | -0.2175 | 0.7733 | 0.0003 |
| **Rainfall** | -0.0711 | -0.1323 | 0.2398 | 0.0282 |
| **Sulphur Dioxide** | -0.0198 | -0.0158 | 0.7439 | 0.7943 |
| **Temperature** | 0.6238 | 0.3582 | 0 | 0 |
| **Wind Direction (10m)** | 0.3363 | 0.0313 | 0 | 0.6056 |
| **Wind Direction Sigma Theta** | 0.1988 | -0.1319 | 0.0009 | 0.0288 |
| **Wind Speed (10m)** | -0.0604 | 0.3787 | 0.3180 | 0 |

Actual and predicted PBL heights in Merriwa vs external factors:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | correlation | | p-value | |
|  | CL51 | CT Model | CL51 | CT Model |
| **Carbon monoxide** | 0.5229 | 0.2004 | 0 | 0.0119 |
| **Humidity** | -0.7558 | -0.5974 | 0 | 0 |
| **Nephelometer** | -0.1288 | -0.0894 | 0.1080 | 0.2653 |
| **Nitric Oxide** | 0.0529 | 0.1509 | 0.5103 | 0.0593 |
| **Nitrogen Dioxide** | -0.1041 | -0.3127 | 0.1943 | 0.0001 |
| **Ozone** | 0.6958 | 0.4978 | 0 | 0 |
| **PM10** | 0.4756 | 0.0912 | 0 | 0.2557 |
| **PM2.5** | 0.0788 | -0.2451 | 0.3264 | 0.0020 |
| **Rainfall** | -0.0654 | -0.1315 | 0.4157 | 0.1008 |
| **Sulphur Dioxide** | 0.0567 | -0.0469 | 0.4807 | 0.5599 |
| **Temperature** | 0.6925 | 0.5277 | 0 | 0 |
| **Wind Direction (10m)** | 0.4001 | 0.1277 | 0 | 0.1111 |
| **Wind Direction Sigma Theta** | 0.0058 | 0.0102 | 0.9424 | 0.8990 |
| **Wind Speed (10m)** | 0.1100 | 0.3993 | 0.1704 | 0 |

Actual and predicted height in Lidcombe vs external factors:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | correlation | | p-value | |
|  | CL51 | CT Model | CL51 | CT Model |
| **Carbon monoxide** | -0.2849 | -0.4609 | 0.0018 | 0 |
| **Humidity** | -0.5770 | -0.4378 | 0 | 0 |
| **Nephelometer** | -0.3133 | -0.3114 | 0.0006 | 0.0006 |
| **Nitric Oxide** | 0.1525 | -0.2497 | 0.0992 | 0.0064 |
| **Nitrogen Dioxide** | -0.0140 | -0.6006 | 0.8804 | 0 |
| **Ozone** | 0.3286 | 0.3991 | 0.0003 | 0 |
| **PM10** | 0.1387 | -0.1190 | 0.1342 | 0.1993 |
| **PM2.5** | -0.1481 | -0.1481 | 0.1095 | 0.1096 |
| **Rainfall** | -0.0889 | -0.1939 | 0.3387 | 0.0354 |
| **Solar radiation** | 0.3891 | 0.6455 | 0 | 0 |
| **Sulphur Dioxide** | 0.0706 | -0.1435 | 0.4474 | 0.1210 |
| **Temperature** | 0.4827 | 0.1878 | 0 | 0.0418 |
| **Wind Direction (10m)** | 0.1866 | -0.1182 | 0.0431 | 0.2023 |
| **Wind Direction Sigma Theta** | 0.2264 | -0.2264 | 0.0137 | 0.0137 |
| **Wind Speed (10m)** | -0.0342 | 0.5401 | 0.7128 | 0 |

1. Obtained from: https://www.environment.nsw.gov.au/topics/air/understanding-air-quality-data/glossary-of-air-quality-terms [↑](#footnote-ref-1)