Predicting Air Quality from Low-Cost Sensor Measurements

Hamish Huggard[1], Yun Sing Koh[1], Patricia Riddle[1], and Gustavo Olivares[2]

[1] The University of Auckland, New Zealand
hamishhuggard@gmail.com,{ykoh, pat}@cs.auckland.ac.nz
[2] National Institute of Water and Atmospheric Research, New Zealand
Gustavo.Olivares@niwa.co.nz

Abstract. Urban air pollution poses a significant global health risk, but due to the high expense of measuring air quality, the amount of available data on pollutant exposure has generally been wanting. In recent years this has motivated the development of several cheap, portable air quality monitoring instruments. However, these instruments also tend to be unreliable, and thus the raw measurements require preprocessing to make accurate predictions of actual air quality conditions, making them an apt target for machine learning techniques. In this paper we use a dataset of measurements from a low cost air-quality instrument—the ODIN-SD—to examine which techniques are most appropriate, and the limitations of such an approach. From theoretical and experimental considerations, we conclude that a robust linear regression over measurements of air quality metrics, as well as relative humidity and temperature measurements produces the model with greatest accuracy. We also discuss issues of concept drift which occur in this context, and quantify how much training data is required to strike the right balance between predictive accuracy and efficient data collection.

Keywords: Air Quality, Polynomial Regression, Concept Drift

1 Introduction

The term particulate matter (PM) describes the aerosols (solid or liquid) suspended in the air, and is a key pollutant in urban areas. PM has been linked to a range of health issues and the World Health Organization has estimated that PM causes 6.4 million years of life lost every year globally [3]. Understanding and quantifying exposure to air pollutants are essential in many human health applications, including risk assessment and accountability evaluations [9].

Due to the high cost of installing and operating accurate air pollution sensors [10], measurements of PM concentrations tend to be sparse in space, and often sparse in time. More precise knowledge of spatiotemporal PM distributions would therefore be of benefit to public health. In recent years, considerable effort has been put in the development of portable, low-cost air pollution monitoring instruments [12]. This reduction in cost allows for more measurements to

be collected over a given region of space, but also means that each individual measurement is less accurate. Determining how to extract useful measurements and uncertainties from the raw measurements of these low cost sensors would therefore be a useful application of machine learning.

To illustrate this type of application, we examine a dataset generated from a pilot study by the National Institute of Water and Atmospheric Research (NIWA). For this study, a fleet of 18 low-cost air pollutant monitoring instruments called Outdoor Dust Information Nodes - Size Distribution (abbreviated ODIN-SD or ODIN) were deployed throughout Rangiora, a town in the Canterbury plains, New Zealand during the Winter of 2016 (illustrated in Figure 1). Due to the unknown accuracy of the ODINs, for stretches of the experiment they were colocated with a TEOM-FDMS (a regulatory grade instrument used for official measurements) at Environment Canterbury's (ECan's) air quality monitoring site in central Rangiora so that the two sets of measurements could be compared. Figure 2 illustrates when each ODIN was at the ECan site. These periods provide us with labeled datasets from which we can build models to predict actual PM concentrations from raw ODIN measurements.

The key contributions of this research are:

– We argue that situations such as this are highly adversarial to machine learning techniques: the data are noisy, multi-dimensional, and subject to both real and virtual drift. It is therefore unreasonable to expect to do better than robust linear regression models, and indeed we demonstrate that such models outperform several more sophisticated models.
– By considering feature selection, seasonal variation, and concept drift detection we are able to recommend a modelling technique for the ODIN-SD instrument. The ODIN has since been deployed to collect data in Idaho, Otago, and Gisborne. Having an appropriate modelling technique at hand will therefore assist with future PM modelling work.
– We estimate the error in PM predictions based on when and how much training data is collected. This should allow more strategic deployment of ODINs in the future for more efficient data collection.

2 The Dataset

The ODIN is equipped with a Plantower PMS3003 dust sensor and a DHT22 temperature and relative humidity sensor. The dust sensor measures three quantities: PM_1, $PM_{2.5}$ and PM_{10}. These denote collections of particles which can pass through a size-selective inlet with a 50% efficiency cut-off at 1, 2.5 and 10 μm aerodynamic diameter, respectively. ODINs sample PM_1, $PM_{2.5}$, PM_{10}, temperature and relative humidity at one minute intervals.

The TEOM-FDMS at the ECan site measures $PM_{2.5}$ and PM_{10} concentrations according to New Zealand's regulatory standard (AS/NZS 3580.9.16:2016[3]). We will refer to these measurements as $PM_{2.5,ECan}$ and $PM_{10,ECan}$, respectively.

---
[3] https://shop.standards.govt.nz/catalog/3580.9.16%3A2016%28AS%7CNZS%29/view
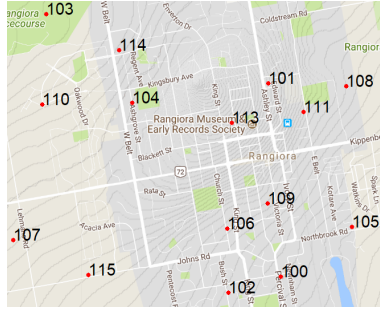
Fig. 1: ODIN deployments in Rangiora overlayed on a Google Map. Locations are labeled by ODIN serial number.
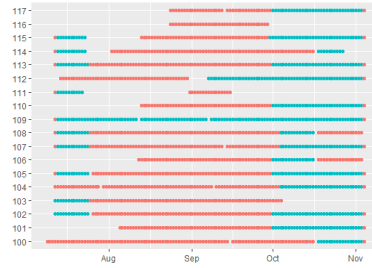

Fig. 2: Timeline of ODIN locations. Blue denotes being located at the ECan site and red denotes deployment.

ECan data was obtained through ECan's Data Catalogue[4]. These measurements are given as hour averages, whereas ODIN measurements are taken at ten minute intervals, so we needed to apply a sliding window average to the ODIN data for them to be comparable. We investigated whether the ODIN and ECan site clocks were well synchronized and found that displacing the timestamps in either direction tended to reduce the correlation between measurements [5]. A policy of not offsetting ODIN time stamps was therefore chosen.

From Figure 2 it is apparent that the data requires sanitisation before useful comparisons may be made between ODINs. It will be useful to extract two sub-datasets from the complete set of measurements. ODIN-109 (that is, the ODIN with serial number 109) is at the ECan site throughout the experiment. We therefore call the set of measurements from ODIN-109 the *Single ODIN Dataset*, which will be useful for evaluating ODIN performance over medium-length time horizons. Several of the ODINs were concurrently at the ODIN site for the first 12 days and the last two weeks of the experiment, so we will call this set of ODINs the *Multi-ODIN dataset*. Because these datasets both deal with reasonably small time spans and quantities of data, we are unable to draw any strong or general conclusions about PM prediction policies. Instead we make some practical recommendations for effective use of the ODIN-SD and similar instruments in future PM monitoring work. The main results of this work are in Section 5, on model selection, and Section 6 on concept drift.

3 Related Work

The importance of modelling air quality has attracted much attention from the machine learning community in recent years, and sophisticated air quality models

---

[4] http://data.ecan.govt.nz/Catalogue/Method?MethodId=94

[5] Our code is available at https://github.com/lajesticvantrashell/ODINs

have been developed. These have generally been designed for the scale of large cities, with spatial resolution at the order of 1 km^2 pixels. These have often also included real-time forecasts [8] and even web interfaces [15].

The philosophy of these approaches has been to use sophisticated models to interpolate between the sparse measurements from high-cost air quality sensors. That is not the goal of this paper. Instead, we are concerned with the case where air quality measurements are spatially dense, but from low-cost, inaccurate sensors. For this application, more rudimentary interpolation techniques may be used, so our concern is with the problem of predicting local PM conditions based on the unreliable raw measurements of the sensors.

The Purple Air PA-I, and its successor the PA-II are low-cost air-quality instruments, which, like the ODIN, use a Plantower PMS dust sensors from earlier and later in the design cycle, respectively. These instruments have undergone both lab and field tests [2]. Like many low-cost PM sensors, these tests demonstrated that the Purple Air instruments can measure PM$_1$ and PM$_{2.5}$ concentrations much more accurately than PM$_{10}$ concentrations. This is extremely likely to also apply to the ODIN, so this paper focuses on predicting PM$_{2.5,\text{ECan}}$ (rather than PM$_{10,\text{ECan}}$) from ODIN measurements, which will probably be more successful. These tests also revealed that at the conjunction of low temperature, high humidity, high PM concentration, the dust sensor accuracy is worst.

4   Accuracy

In this section we explore two questions relating to the accuracy of PM predictions. The first is how much training data do you need to collect to achieve a given level of PM prediction accuracy. The second is whether it matters when in the season the training data is collected.

How much training data must be collected? For each ODIN a separate set of training data is required. We therefore have a trade off: the more training data collected per ODIN, the more accurate the model will be. However, more time collecting training data means less time collecting new and useful measurements. In this section we empirically quantify this trade off.

To estimate the error of a model trained on n days worth of data (where $n \in \{1, 2, \ldots, 40\}$, a plausible range of colocation durations), we divided the Single ODIN Dataset into noon-to-noon 24-hour periods $(D_1, \ldots, D_n)$. Then for each $i \in \{1, 2, \ldots, 30\}$, we created a training dataset of n days data, starting with D_i: $Train_{i,n} := \{D_i, D_{i+1}, \ldots, D_{i+n-1}\}$, and a test set of the following month's data: $Test_{i,n} := \{D_{i+n}, \ldots, D_{i+n+29}\}$. We could then build a model from $Train_{i,n}$, and the MSE calculated using $Test_{i,n}$ (this value is then called Err$_{i,n}$). We then estimated the error of a model trained on n days data as

$$\text{Err}_n = \frac{1}{30} \sum_{i=1}^{30} \text{Err}_{i,n} \tag{1}$$

The purpose of averaging over a range of i values is to account for the fact that model accuracy may depend on the dates of when training data is collected. By starting the training data period on different dates we may achieve an error estimation that applies to a broad range of training start-dates.

This is plotted in Figure 3. We see that for up to 40 days of training data, the MSE of a model decreases approximately linearly. Performing a linear regression allows us to estimate the MSE for a model trained on $n \in \{1, \ldots, 40\}$ days' colocation data as $\mathrm{Err}_n = (44.48 \pm 0.53) - (0.23 \pm 0.02) \times n$. After 35 days the MSE actually increases when more days' training data are added. This may be because only training windows with at least 35 days will include the anomalous period seen in Figure 4.

Seasonal Dependencies of Accuracy In addition to understanding model sensitivities to the amount of training data available, it is also useful to check for sensitivities to when in the season the training data is collected. The distribution of PM concentrations changes over the course of the winter, and different distributions of training data may yield better models than others. To explore this, we again tested the performance of linear regression models on the Single ODIN Dataset. We iterated through every noon-to-noon week-long and month-long window, used the data from this window as a training set and the rest of the data as a test set. The MSE of these models are illustrated in Figure 4.

As before, we performed a simple linear regression over these error estimates, this time using the date on which the training data was collected as the regressor. This revealed positive trend between the MSE and how late in the season the training data is collected. Each day's delay in the collection of training data added an average of 0.35 ± 0.20 (95% CI) to the MSE of a model trained on one week's worth of data, and added an average of 0.35 ± 0.06 (95% CI) to the MSE of a model trained on one month's worth of data. In either case, there is a statistically significant positive relationship between how late in the season the training data is taken and the accuracy of the model ($p < 0.05$). By examining Figure 4, we see that there is a period near the end of September when the error is anomalously high. This could have something to do with daylight saving time starting (this occurred on the 25th of September), when human and natural daily cycles deviate from one another. Whatever the cause, it is not at all clear that the positive trend between lateness of training data and accuracy of resulting models exists beyond the fact that when training data is taken later in the season, it is more likely to coincide with the anomalous period. The only practical policy to be gleaned from this result seems therefore to be to avoid collecting training data during the end of September. Whether better results are obtained by taking training data from early in Winter is unclear.

5 Model Selection

In this section we attempt to find a suitable model for predicting $\mathrm{PM}_{2.5,\mathrm{ECan}}$ from ODIN measurements.
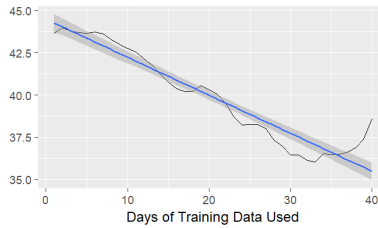
Fig. 3: Estimation of the MSE of a model dependent on the number of days training data used.
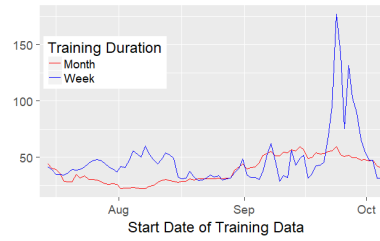

Fig. 4: Estimation of the absolute MSE of a model versus when in the season the training data is collected (x-axis).

Bias Variance Trade-off Choosing a model involves a trade-off between bias and variance. Bias, or approximation bias, is error resulting from poor choice of model, and variance, or estimation variance, is error resulting from the model being trained on some particular training set, rather than on the actual underlying distribution.

There are two main classes of models available: parametric and nonparametric. For nonparametric models – including kernel regression, k-nearest neighbour, splines, and local polynomials – the MSE of the model follows

$$MSE_{nonpara} = \sigma^2 + O(n^{-4/(p+4)}) \tag{2}$$

where the σ^2 term is the intrinsic noise term, $O(n^{-4/(p+4)})$ is the estimation variance term, n is the size of the training set and p is the dimensionality of the feature vector [11]. Here $p = 3$, so the variance term becomes $O(n^{-4/7})$.

Compare this with the MSE for a (nonparametric) linear model:

$$MSE_{linear} = \sigma^2 + a_{linear} + O(n^{-1}). \tag{3}$$

The estimation variance term shrinks faster with more training data than in the nonparametric case. However, the additional term a_{linear} represents approximation bias resulting from the "true" relationship deviating from linearity.

Figure 5 shows that the relationship between $PM_{2.5,Ecan}$ and $PM_{2.5,ODIN}$ is approximately linear, although it is unlikely that it is exactly linear. Previous attempts to apply a linear model to Plantower PMS dust sensor measurements have been fairly successful [2], indicating that the approximation bias of a linear model will be small. On the other hand, Figure 5 also reveals that the data is very noisy, so estimation variance will probably be large. From these considerations, it seems very likely that models with fewer degrees of freedom will outperform more complicated models, even if these other models have the capacity to more closely capture the true relationship between the variables.

In particular, it seems very likely that parametric models will perform better than nonparametric ones (a_{linear} will be small compared to the difference between the $O(n^{-1})$ and $O(n^{-4/7})$ terms from Equations 2 and 3). Nonparametric

Table 1: Results of 2-fold cross-validation of models using data from initial collocation of several ODINs.

	Linear MSE	Quadratic MSE	Cubic MSE	Reg. Tree MSE
Serial				
102	52.14	56.47	59.9	98.21
103	39.59	40.37	42.93	82.59
105	100.23	131.95	134.80	129.66
107	74.30	77.51	79.93	94.29
108	52.99	60.93	68.17	98.06
109	50.71	48.45	43.99	74.58
111	64.70	72.98	80.04	86.23
113	88.07	89.81	85.50	108.09
114	71.71	63.21	75.95	81.59
115	38.31	39.25	40.79	78.37
Average	63.27	68.09	71.2	93.17
Std.	20.40	27.61	27.94	16.54

Table 2: Results of 3-fold cross validation of models using ODIN-109 data.

Model	MSE
Linear	35.31
Quadratic	37.80
Cubic	155.65
Reg. Tree	51.61

models generally require fine tuning of hyperparameters (for example, kernel regression requires bandwidth selection for each dimension), so they take considerably more effort to implement than parametric models. We therefore tried several parametric models before trying any nonparametric ones, to determine whether increasing the degrees of freedom yields greater accuracy.

From the above considerations, we chose to test additive polynomials of up to order three. That is, a linear, quadratic and cubic model. In addition, we also tested a regression tree model, to make sure there were no large gains which could be easily achieved from a nonparametric model, even if it is not well-suited to the modeling task.

Multi-ODIN Dataset The standard method of error validation in situations such as these is k-fold cross validation. Because this data is a time sequence, data points close together in time will be correlated, and so we compose folds of contiguous blocks of data to avoid overfitting.

The results of performing a 2-fold cross validation on the Multi-ODIN Dataset with several models are given in Table 2, along with the inter-ODIN mean and sample standard deviation. We see that no model significantly outperforms linear regression, supporting the notion that estimation variance dominates over approximation bias, so out of parsimonious considerations we should use a linear model.

Single ODIN Dataset The above experiment validated the use of a linear model across ODINs over a short time horizon. We next performed a 3-fold cross validation on the Single ODIN Dataset to verify that the model holds up over a medium time horizon. The results are given in Table 2. Once again, the
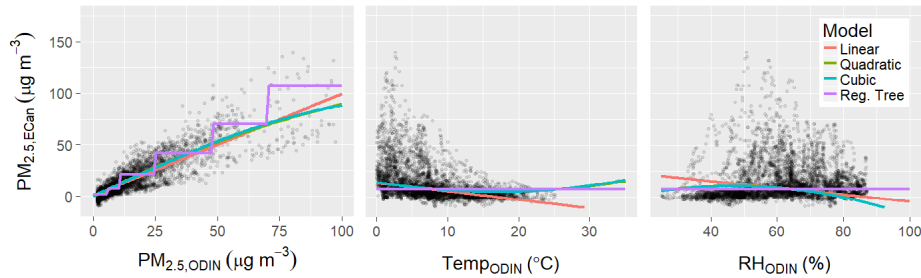
Fig. 5: On the x-axes are the measurements made by the ODIN, compared to the actual $PM_{2.5}$ values (as measured by the ECan TEOM-FDMS) on the y-axis. These data points come from the first third of the Single ODIN dataset, and the lines designate the results of different modelling techniques.

linear model is not significantly outperformed. From these two experiments, we can have fairly high confidence that the linear model is the best choice, and especially that investing more effort into developing a more complex model will not pay off with a significant improvement in accuracy. Interestingly, consistent across all models, the mean error from using the third fold as the training data is significantly larger those obtained with the other two folds. Figure 4 sheds some light on this, by demonstrating that there is an anomalous region during the latter third of the experiment.

Robustness To extend the above result, we would also like to test whether a robust linear regression model performs better than a simple linear regression model. Given that the noise is high enough for estimation variance to dominate over approximation bias, it seems likely that greater robustness against outliers would improve performance. We performed a 3-fold cross validation using an M estimator on both the Single and Multi-ODIN Datasets. For the Multi ODIN Dataset, the mean square error estimation dropped from 63.27 ± 20.40 to 59.06 ± 18.13 when switching from linear to robust linear (these are the inter-ODIN averages, with standard deviations) and in the Single ODIN Dataset the mean square error estimation dropped from 35.31 to 34.86. Although these reductions are not significant, adding robustness *a priori* seems sensible for noisy data such as these. We therefore believe that the best model in this context is produced by robust linear regression.

Feature Selection Work on a previous iteration of an ODIN found that the dust sensor's performance varied with temperature and relative humidity conditions [10]. This dataset is small enough that we were able to apply a brute-force feature selection technique to verify that using a feature vector of $(PM_{2.5,ODIN}, Temp_{ODIN}, RH_{2.5,ODIN})$ did indeed yield a more accurate model than using $PM_{2.5,ODIN}$ as a sole regressors.

6 Concept Drift

In this section we attempt to determine whether the ODIN measurements are stable over time, or if sensor degradation is occurring. In the study on the earlier ODIN, it was discovered that significant baseline drift occurs over the course of several weeks, rendering a baseline correction necessary [10]. We do not know if such a correction is required in this case. If sensor degradation *is* occurring, then this would be concept drift, *i.e.*, a non-stationary learning environment.

Virtual and Real Concept Drift It will be useful here to draw on the distinction sometimes made between real and virtual concept drift: real drift is a physical change in the actual learning environment, whereas virtual drift does not occur in reality, but only in the computer model [14]. This may be a result of a changing context altering the distribution of the data, and thus affecting the model error in such a way that requires the model to adapt. The distribution of $PM_{2.5}$ is changing over time: earlier in the season, there are many more instances of high $PM_{2.5}$ concentrations. This makes it seem likely that virtual drift will be occurring, so it will be difficult to differentiate this from real drift.

There are plenty of techniques for detecting concept drift, either real or virtual. For example, one can monitor the raw data stream [1], the error rate of the learner [5] or the parameters of the model itself [13]. However, most of these techniques are designed for classification, rather than regression problems. Furthermore, the problem of detecting real drift apart from, but in the presence of, virtual drift has not been previously explored in any depth.

Evolving models which adapt to concept drift - whether via adaptive ensembles [6], instance weighting [7] or adaptive sampling [4] may perform well under these conditions. However, because these techniques require the receipt of correct labels after they have made their predictions, they are useless for our purposes: once an ODIN is deployed, we will not have access to "correct" labels.

Detection of Real Drift Fortunately, in this particular case we *do* have a way of detecting real drift in the presence of virtual drift. Many of the ODINs were at the ECan site at both the beginning and the end of the experiment. For each of these ODINs we can therefore train models on both the initial and final ECan collocations. If the parameters change from the first case to the second, we cannot determine whether this is due to real or due to virtual drift. However, if we have data from two ODINs from the same time intervals, then because these data are collected from identical PM conditions any virtual drift which occurs should result in parameters moving in the same direction for both ODINs. Thus if the model parameters drift apart, then we have evidence of real drift (although we will not be able to tell which of the ODINs are degrading, nor in which direction).

To implement this idea, we used data from ODINs 102, 105, 107, 108, 109, 113 and 115, which have a reasonably large intersection of colocation times. We then performed linear regressions over the initial and final colocation datasets

Table 3: The absolute changes of each of the model parameters between being trained on the initial and final collocations with the ECan site. "Cnst." denotes the constant term of the model, and "ΔCoef." denotes the change in coefficient for that measurement between in initial and final collocation. The averages include the standard deviation across ODINs.

Serial	ΔCnst.	$PM_{2.5,ODIN}$ ΔCoef.	$Temp_{ODIN}$ ΔCoef.	RH_{ODIN} ΔCoef.
102	-6.53	-0.01	0.44	0.05
105	0.59	-0.23	0.21	-0.01
107	-6.39	0.29	0.53	0.05
108	-7.35	-0.17	0.33	0.08
109	-1.54	-0.21	0.34	-0.01
113	-14.25	0.15	0.45	0.15
115	-8.21	-0.20	0.31	0.10
Average	-6.24±4.80	-0.06±0.21	0.37±0.11	0.06±0.06

and compared the parameters for the initial and final colocation datasets. Random noise could still cause some inter-ODIN differences in model changes, and to mitigate this effect, we again used `rlm` for robustness against outliers. The changes in the $PM_{2.5,ODIN}$ coefficients are illustrated in Figure 6 (including 95% confidence intervals, estimated as twice the standard deviation given in the `rlm` output), and the changes in all the model parameters are given in Table 3.

We see that the ODIN model parameters are indeed drifting apart from one another, providing good evidence that real drift (*i.e.*, sensor degradation) is occurring. This is especially evident in the case of $PM_{2.5,ODIN}$, which is the feature we are most concerned with. The $PM_{2.5,ODIN}$ coefficients of ODIN-107, ODIN-108, ODIN-109 and ODIN-115 all decrease from the initial to the final colocation by a fairly consistent amount, whereas that of ODIN-105 *increases* by a similar amount and those of ODIN-102 and ODIN-113 do not vary much. Because the $PM_{2.5,ODIN}$ coefficients which do change, change by several confidence intervals, we can be fairly certain that different ODINs are genuinely degrading, and thus that real drift is in fact occurring.

Model Evolution Over Time Having established that some kind of sensor degradation (real drift) is occurring, we now look at the entire ODIN-109 dataset, to try and discern trends in how the model changes over time. By dividing the dataset into 24 hour period (noon to noon) blocks and training a robust linear model on each block, the movements of each model parameter can be plotted over time, as in Figure 7. We notice that the $PM_{2.5,ODIN}$ coefficient is steadily decreasing over time. However, we do not know if this represents real or virtual drift. We can attempt to differentiate the two types of drift by making a correction for the PM conditions of each day.

This can be achieved by performing a simple linear regression over the mean $PM_{2.5,ECan}$ value and the date of each 24-hour period, using the $PM_{2.5,ODIN}$ co-
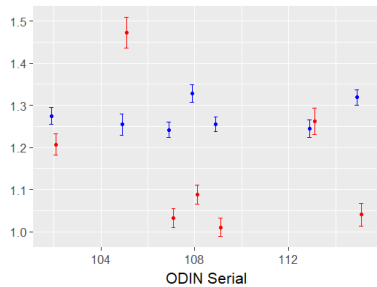
Fig. 6: $PM_{2.5,ODIN}$ Coefficient of robust linear regression model built using initial (blue) and final (red) collocation data.
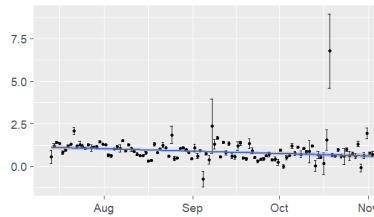

Fig. 7: The $PM_{2.5,ODIN}$ coefficients of linear regression models trained on successive noon-to-noon 24 hour periods of ODIN-109 data.

efficient as the dependent variable. The resulting time-dependent slope indicated that the $PM_{2.5,ECan}$ coefficient on average decreases by $(1.08 \pm 5.46) \times 10^{-3}$ (95% CI) per day, regardless of actual PM conditions for that day. This indicates that real drift is occurring in ODIN-109, and that the sensor is degrading in a steady, linear manner over time. Further research is required to characterise this drift more precisely so that drift corrections can be incorporated into the model.

7 Conclusion

We examined the question of how best to extract useful information about PM concentrations from an ODIN instrument. We produced a formula relating the amount of training data collected (*i.e.*, the length of the colocation with the ECan site) to the error of the resulting model. This allows more informed decisions about trade offs between the quantity and the quality of the data which can be collected. Whether model accuracy has a dependency on when in the season training data is collected was ambiguous. Because the measurements taken by ODINs are so noisy, estimation variance dominates over approximation bias on the time scales we are interested in. This means that linear regression models yield optimal performance. From there, removing outliers with M estimators likely further improves performance. In particular, feature vectors of $(PM_{2.5,ODIN}, Temp_{ODIN}, RH_{2.5,ODIN})$ seem to work best for linear regression models. We also established that some form of sensor degradation was occurring over the course of the experiment, and attempted to quantify this. However, more work will need to be done on this before useful corrections can be made.

References

1. Albert Bifet and Ricard Gavalda. Learning from time-changing data with adaptive windowing. In *Proceedings of the 2007 SIAM International Conference on Data Mining*, pages 443–448. SIAM, 2007.

2. Air Quality Sensor Performance Evaluation Center. Purpleair PA-ii - summary report. http://www.aqmd.gov/docs/default-source/aq-spec/summary/purpleair-pa-ii—summary-report.pdf?sfvrsn=4. Accessed: 2018-02-20.

3. Aaron J Cohen, H Ross Anderson, Bart Ostro, Kiran Dev Pandey, Michal Krzyzanowski, Nino Künzli, Kersten Gutschmidt, Arden Pope, Isabelle Romieu, Jonathan M Samet, et al. The global burden of disease due to outdoor air pollution. *Journal of Toxicology and Environmental Health, Part A*, 68(13-14):1301–1307, 2005.

4. Sarah Jane Delany, Pádraig Cunningham, Alexey Tsymbal, and Lorcan Coyle. A case-based technique for tracking concept drift in spam filtering. *Knowledge-based systems*, 18(4-5):187–195, 2005.

5. João Gama, Pedro Medas, Gladys Castillo, and Pedro Rodrigues. Learning with drift detection. In *Brazilian Symposium on Artificial Intelligence*, pages 286–295. Springer, 2004.

6. J Zico Kolter and Marcus A Maloof. Dynamic weighted majority: An ensemble method for drifting concepts. *Journal of Machine Learning Research*, 8(Dec):2755–2790, 2007.

7. Ivan Koychev. Gradual forgetting for adaptation to concept drift. Proceedings of ECAI 2000 Workshop on Current Issues in Spatio-Temporal Reasoning,, 2000.

8. Xiaorong Lu, Yang Wang, Liusheng Huang, Wei Yang, and Yao Shen. Temporal-spatial aggregated urban air quality inference with heterogeneous big data. In *International Conference on Wireless Algorithms, Systems, and Applications*, pages 414–426. Springer, 2016.

9. Thomas E McKone, P Barry Ryan, and Halûk Özkaynak. Exposure information in environmental health research: Current opportunities and future directions for particulate matter, ozone, and toxic air pollutants. *Journal of Exposure Science and Environmental Epidemiology*, 19(1):30, 2009.

10. G Olivares and S Edwards. The outdoor dust information node (ODIN)–development and performance assessment of a low cost ambient dust sensor. *Atmospheric Measurement Techniques Discussions*, 8:7511–7533, 2015.

11. Cosma Shalizi. *Advanced data analysis from an elementary point of view*. Cambridge University Press Cambridge, 2013.

12. Emily G Snyder, Timothy H Watkins, Paul A Solomon, Eben D Thoma, Ronald W Williams, Gayle SW Hagler, David Shelow, David A Hindin, Vasu J Kilaru, and Peter W Preuss. The changing paradigm of air pollution monitoring, 2013.

13. Bai Su, Yi-Dong Shen, and Wei Xu. Modeling concept drift from the perspective of classifiers. In *Cybernetics and Intelligent Systems, 2008 IEEE Conference on*, pages 1055–1060. IEEE, 2008.

14. Gerhard Widmer and Miroslav Kubat. Effective learning in dynamic environments by explicit context tracking. In *European Conference on Machine Learning*, pages 227–243. Springer, 1993.

15. Yu Zheng, Xiuwen Yi, Ming Li, Ruiyuan Li, Zhangqing Shan, Eric Chang, and Tianrui Li. Forecasting fine-grained air quality based on big data. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 2267–2276. ACM, 2015.