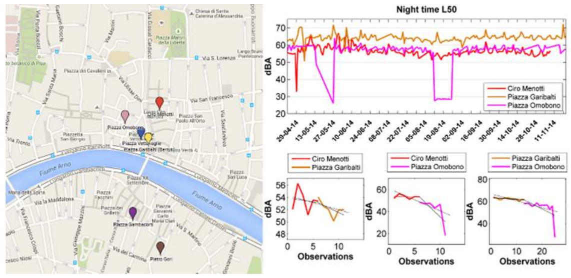
A Statistical Method for Assessing Network Stability Using the Chow Test

Journal:	*Environmental Science: Processes & Impacts*
Manuscript ID:	EM-ART-07-2015-000325.R1
Article Type:	Paper
Date Submitted by the Author:	02-Sep-2015
Complete List of Authors:	Sotirakopoulos, Kostas; National Physical Laboratory, Acoustics and Ionizing Radiation Barham, Richard; National Physical Laboratory, Acoustics and Ionizing Radiation Piper, Ben; National Physical Laboratory, Acoustics and Ionizing Radiation Nencini, Luca; Bluewave acoustics, Research and development

A computationally inexpensive statistical method for the identification of drifts from calibration in noise monitoring wireless sensor networks.



Environmental Impact Statement

In modern cities the need for monitoring of environmental influences on peoples' wellbeing and behaviour has become crucial. Noise is undoubtedly a very important environmental parameter that has an immediate impact on populations residing in urban environments. Technological advances in wireless sensor networks have made possible the creation of low cost noise monitoring networks that can provide alternatives to traditional, expensive noise monitoring applications. The resulting information can be used to inform the public and assist the decision making of urban planners, hence it is important that there is no compromise on its quality. This paper presents a method for assessing stability of such networks in order to ensure high measurement quality to facilitate decision making based on accurate data.

Journal Name

ARTICLE

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

A Statistical Method for Assessing Network Stability Using the Chow Test

Kostas Sotirakopoulos, Richard Barham, Ben Piper, Luca Nencini

A statistical method is proposed for the assessment of stability in noise monitoring networks. The technique makes use of a variation of the Chow Test applied between multiple measurement nodes placed at different locations and its novelty lies in the way it utilises a simple statistical test based on Linear Regression to uncover complex issues that can be difficult to expose otherwise. Measurements collected by a noise monitoring network deployed in the center of Pisa are used to demonstrate the capabilities and limitations of the test. It is shown that even in urban environments, where great soundscape variations are exhibited, accurate and robust results can be produced regardless of the proximity of the compared sensors as long as they are located in acoustically similar environments. Also it is shown that variations of the same method can be applied for self-testing on data collected by single stations. Finally it is presented that the versatility of the test makes it suitable for detection of various types of issues that can occur in real life network implementations; from slow drifts away from calibration, to severe, abrupt failures and noise floor shifts.

Journal Name

ARTICLE

Contents

Journal Name

ARTICLE

Introduction

As societies turn from traditional structures to modern centralized organization models, big cities become home to a continuously growing number of people. Over the past decades, huge populations migrated from rural to urban environments and this trend only becomes more profound with time. It is a fact that already 70% of EU's population resides in urban areas while studies predict an increase in worldwide cities' demographics from 29% in 1950 to a remarkable 65% by 2040 [1]. From Japan to Mexico one can find examples of Mega Cities housing tens of millions and given the demographic predictions it is only a matter of time before their number multiplies. As a result, the issue of improving and preserving life quality in such environments becomes of great importance. Modern technologies and monitoring networks in particular can assist communities in continuously monitoring and sharing information related to a city's environment and efficiency; hence taking a step forward to transforming big cities into Smart Cities.

Noise monitoring networks is a concept that has been developed over the last years making use of technological advances in sound measuring instrumentation. Various approaches have been adapted so far; from dynamic grids using smart phones [2] to the creation of more sophisticated stationary monitoring stations. The results of such studies show that smart phones, although convenient as measuring devices, for a number of reasons are not suitable for delivering trustworthy measurements [3] [4]. On the other hand the development of components such as MEMS (Micro-Electro-Mechanical-Systems) microphones and their integration into monitoring stations capable of being installed in large numbers due to their low manufacturing cost has proved to be a very promising perspective [5]. At the National Physical Laboratory in the U.K. research has focused on the development of improved quality yet cheap MEMS microphones which will fit such purposes and their capabilities have been explored on various applications [6] [7] [8] [9] [10].

While significant effort has been put into achieving low cost implementations of monitoring systems there is still an additional cost involved with the installation of grids consisting of multiple measurement stations. That is expenses related to the maintenance and quality control of the deployed network making it clear that the development of automated methods for evaluating network stability and thus measurement quality is crucial. A great deal of work has been done on the identification of anomalies in data collected by distributed networks; V. Chandola et al [11] and Y. Zhang et al [12] provide comprehensive surveys on existing techniques ranging from statistical to rule based and machine learning methods.

Some approaches make use of the concept of combining multiple criteria in order to identify different types of anomalies. S. Dauwe et al [13] demonstrate how the weighted sum of four, or even more, tests including laboratory measurements, comparisons of sound events between various time intervals as well as more complicated models based on the simulation of the human auditory system, can provide a single index for network integrity evaluation.

Other anomaly detection systems rely on the comparison of sensor measurements to closely positioned monitoring stations to derive calibration functions for each measurement node [14] [15] . However such approaches often require dense networks which in real life applications might prove expensive to implement. Similar techniques employ

Data Fusion Models for networks organized into clusters in order to detect faults on individual sensors and calculate the calibration coefficients for each sensor [16].

A common anomaly observed in data collected by low cost sensor networks is slow deviation from calibration which in some cases happens over months or even years of exposure. It is exactly this slow nature of its occurrence that makes it rather difficult to detect leading to long periods of faulty data collection. In this paper we present a method which utilises a linear regression based statistical test applied on statistical noise levels in order to address this issue. This method falls in the greater category of techniques comparing routine measurements between multiple nodes, while it can perform well without requiring oversampling of the soundscape. It is also shown that this technique can be used to detect other much more severe anomalies like system failures. Examples are demonstrated on experimental data collected by a noise monitoring network deployed in the city of Pisa in Italy.

Concept

The method discussed is an alternative approach to error detection (that doesn't aim to completely replace traditional calibration), and the presented research is to test this alternative concept. The underlying idea behind the technique stems from the fundamental property of linearity and time invariance (LTI), which any piece of measurement instrumentation must exhibit under normal working conditions. This implies that over time the general image of data collected by any stable sensor must be distributed around a straight line, with some variance due to environmental characteristics which of course change with time and season. Unfortunately, low cost networks exposed to a number of destructive factors that can lead to physical sensor degradation, like varying weather conditions and exposure to public, might not behave like true LTI systems.

Physical fatigue or system failures cause deviation from normal operation which in collected data shows as deviation from the straight line around which 'normal' data exist. Depending on the severity of the damage in the system the magnitude of such deviations varies. Abrupt, severe faults will show as discontinuities while slow drifts from calibration will most probably appear as downward or upward trends in the system's response. In both cases such behaviours could be exposed by comparison with other sensors in the same network given that it is highly improbable for multiple sensors to exhibit identical faulty behaviour at the exact same time.

Figure 1 presents an example of daily averaged noise time series collected between September 2013 and February 2015 by two monitoring stations in the city of Pisa, in Italy. These monitoring stations normally measured one second, broadband L_{eq} levels. However since drifts from calibration usually evolve during long time periods it was considered that daily averages would reduce the variance in the data and at the same time provide a good enough resolution in order to observe such phenomena. Focusing on obvious events, a two weeks long system failure during August 2014 is observed for the sensor placed at Piazza Omobono while a clear drift is seen for the same sensor from approximately November 2014 until the end of the presented dataset. Closer inspection will reveal that this drift may actually begin about a year earlier, at a much slower rate though. However it is masked by its soft nature and the variance in the data. At this point it should be noted that the drift mentioned has most probably occurred due to physical degradation of the microphone caused by insufficient water proofing of the specific unit. On the contrary, no such trend is seen in the data collected at Piazza Vettovaglie. Finally, the observed dip occurring at both time series between June and September 2014 (ignoring the system failure during August for Piazza Omobono) might be mistaken for an anomaly if they are examined separately. Nevertheless, a comparison between multiple sensors suggests that this is a seasonal feature which appears during the summer when the population density of the city decreases.

From the simple examination described above it becomes clear that the combined information sourced by multiple sensors in a measurement network can provide useful knowledge about the network's stability. In particular, if the variance in the data is reduced then linear regression can be a useful tool in revealing anomalies.

The two questions rising from the above thought are:

1. Is there an indicator that presents adequately low variance?
2. Which would be a reasonable way of comparing regressions in order to expose deviations from normal operation?

Simple routine measurements provide the answer to the first question. No matter how variant the soundscape of a city might be, locally the median will exhibit less variation with time. These level variations are further reduced during the late night (3:00 am - 7:00 am). Figure 2 presents these night time L_{50} levels measured at Piazza Vettovaglie and Piazza Omobono. It is obvious that the variance in the L_{50} data is much smaller when compared to the Leq data shown in Figure 1 for the same time period, making the existing trends much more pronounced. For this reason in the analysis presented in the following sections L_{50} night time refers to L_{50} calculated over the time interval mentioned above. If required, further reduction of the variance could be achieved by exclusion of measurements held over the weekends when the acoustic environment becomes quieter mostly due to lower traffic noise levels. Additionally, it is considered that since most of the noise energy measured in urban environments is concentrated at lower frequencies appropriate filtering can aid towards making possible deviations from normal operation more easily detectable.

The Chow Test

Gregory Chow, a Chinese American economist, in 1960 suggested a statistical method for analysing a set of the same variables obtained in different time periods in order to evaluate their similarity and decide whether they can be pooled together [17]. The technique was named The Chow Test, and it is well-known in economics and econometrics where it is mostly used to check for structural breaks in time series. Over time, significant research has been done on variations and the method has found use in various applications [18] [19] [20] [21] [22] [23].

The purpose of the test is to conclude on the significance of the difference between coefficients of two regression lines applied on two data sets. Originally the method would test observations taken from within the same population at different time periods to examine if the relationship between dependent and independent variables changes with time. However this can be generalised and applied to situations where the data come from different populations.

In order to perform the test on two samples consisting of n and m observations respectively, one must start with the null hypothesis, H_0, that the coefficients of the two regressions are equal. Regression is applied on each data set separately and the sum of squared residuals (SSR) is computed. Then the data are combined sequentially to form a $n + m$ long set of observations and a regression line is fitted on it while the sum of squared residuals is computed again. In his work, Chow proved that the ratio of the difference between the combined and individual SSR over the sum of the individual SSR scaled for the corresponding degrees of freedom will follow an F distribution under the null hypothesis. In the original paper it is demonstrated that when one of the two samples has less observations than estimated regression coefficients then a different F-ratio must be computed than when both samples have sizes greater than the estimated regression weights. Here we are only interested in the latter case.

The application of the test as described above is demonstrated in the following steps:

1. Fit a regression line on the first set of n observations and compute SSR_1

2. Fit a regression line on the second set of m observations and compute SSR_2

3. Apply linear regression on the combined set of $n+m$ observations and compute SSR_3

4. The test of the null hypothesis that the second set of data follows the same trend as the first is given by

$$F = \frac{(SSR_3 - (SSR_1 + SSR_2))/p}{(SSR_1 + SSR_2)/(n+m-p)} \quad (1)$$

where p is the number of regression coefficients. This ratio follows an F distribution with $n+m-2p$ degrees of freedom.

Practical Implementation of the Test for Identification of Errors in Noise Measurement Networks

The Chow test, as described so far, is almost guaranteed to lead to rejection of the null hypothesis when applied on time series of noise data collected at different locations in a city. This is due to the fact that the test is meant to examine equality between sets of regression coefficients. However when linear regression is applied there are two coefficients to be estimated; the slope and the offset of the regression line. In urban environments sound levels can vary quite significantly from one place to the other. This in turns means that, as seen in Figure 2, it is very likely to observe significant differences between noise levels recorded at different locations. Hence, the only way to apply the test in its original form between different monitoring systems would be to include redundant (duplicate) sensors at each location and run comparisons between measurement and reference units. Nevertheless, something like this would increase the cost of the application significantly.

In order to overcome this issue an intermediate step should be added before combining data sets coming from different locations. That is scaling of one of the two sets in order to minimize discontinuities in the combined data. To achieve that, the difference between the level at the last point of the regression line fitted on the first data set and the level of the first point of the fitted line on the second data set is computed and used as a scaling factor. This effectively eliminates any discontinuities between the regression lines and turns the focus of the test on their slopes. An example of combined L_{50} data before and after scaling is demonstrated in Figure 3. The data were collected between 20 June and 20 July 2014 at Piazza Gambacorti and Piazza Vettovaglie. For the whole measurement period no particular errors were detected for any of the two nodes, as seen in Figure 4.The critical value above which the null hypothesis is rejected is $F_{crit}(2{,}58) = 1.65$ at $a = 0.2$ level of significance. The reason for selecting $a = 0.2$ as level of significance was because it proved to provide an acceptable threshold for the trade-off between Type 1 and Type 2 errors as discussed in [24].

When no scaling was applied the F score was $F = 52.39$ indicating, incorrectly, a statistically significant error in the system. After scaling however the F score was computed equal to $F = 0.12$ showing a good agreement between the two data sets. Note that the downward trend in the tested data is due to decreasing noise levels during the first months of the summer which return back to normal by the end of September. The test however is not affected by such seasonal characteristics.

The procedure described above is necessary when samples from different locations are compared. However there is one occasion when scaling will have a destructive effect on the validity of the results. Consider a sensor that operates normally until time t_1 when it starts drifting from calibration and continues drifting until some time $t_2$when it stabilizes again. Using the Chow Test for multiple sensors after t_2 will not detect any anomaly as it will simply correct for the level difference between sensors and as long as the slope of the compared time series is similar then it will give a good agreement between regression lines. Such errors are very unlikely to happen since it is rather

Please do not adjust margins

uncommon for a sensor to break down for some period and then stabilize, unless of course it reaches noise floor in which case application of the test would not be required to identify the error. However, there is still a way to spot such anomalies. Under the assumption that there is some annual periodicity in noise levels collected locally (unless of course major changes take place) one could skip the scaling step and apply the original version of the test on data collected over same periods but different years at only one location. This way possible faults like the one described in this paragraph will appear as offset differences between the individual regression lines and will give high F ratios just like in the example depicted in Figure 3.1. Such a test, however, should be used with caution as seasonal characteristics can interfere with the validity of the results.

Application of the Method

In this section a demonstration of the application of the test on various data sets collected during 'Sensable Pisa' project are demonstrated and discussed. The measurement equipment used consisted of a data acquisition board with some processing capabilities, a commercial ¼ - inch condenser low cost microphone while the data transmission to the remote server where data were being stored was achieved using ZigBee protocol. The approximate noise floor of the units was 30dBA. More information about the measurement devices can be found here [25] while a map of the area and exact locations of the monitoring stations is provided in Figure 5. The approximate distances between locations are presented in Table 1.

Table 1: Distances (in meters) between nodes.

Largo Ciro Menoti – Piazza Omobono	150
Largo Ciro Menoti – Piazza Garibalti	150
Largo Ciro Menoti – Piazza Vettovaglie	140
Largo Ciro Menoti – Piazza Gambacorti	450
Largo Ciro Menoti – Via Pietro Gori	570
Piazza Omobono – Piazza Garibalti	130
Piazza Omobono - Piazza Vettovaglie	90
Piazza Omobono - Piazza Gambacorti	380
Piazza Omobono - Via Pietro Gori	530
Piazza Garibalti - Piazza Vettovaglie	40
Piazza Garibalti - Piazza Gambacorti	300
Piazza Garibalti - Via Pietro Gori	430
Piazza Vettovaglie - Piazza Gambacorti	320
Piazza Vettovaglie- Via Pietro Gori	455
Piazza Gambacorti – Via Pietro Gori	175

Case 1: Proximity between measurement nodes

Assuming uniformity of the sound environment between near-by locations, noise readings collected by monitoring stations located at close proximity should demonstrate relatively high correlations. This should always result in good agreement of the regression lines when no system faults occur. However, in urban environments the acoustic characteristics of the soundscape can vary significantly from one location to the other, even over relatively short distances. The above makes the application of the method on multiple nodes positioned at various distances and the comparison of the results an interesting assessment for the robustness of the test when used in urban areas. To demonstrate this, a comparison between night time L_{50} levels measured at Piazza Garibalti , Piazza Vettovaglie, Largo Ciro Menoti and Piazza Gambacorti in the center of the city was applied. As seen in Table 1 all distances, apart from Piazza Garibalti - Piazza Vettovaglie which is only 40 m, can be considered reasonably long for urban environments.

In Figure 6 a drift is seen in the levels measured at Ciro Menotti starting approximately on 12 October 2014 and continuing for about two months until the end of the deployment. However no such trend is met in the data collected at the other three locations. During the deployment period no major alteration in the surrounding environment that could have caused this drift was reported. Moreover the slope of the drift indicates that it could not be caused by a sudden termination of operation of a nearby noise source such as an air conditioning unit or a night pub in the area since this would most probably cause a much more rapid and permanent decrease in the observed levels. On the contrary what is seen is a gradual decrease that lasts several weeks indicating that something else must have happened, possibly an error in the measurement system. The test was run between 29 September and 6 November 2014 to check whether this feature could be identified by all nodes. Figure 7 presents comparisons between fitted lines on individual and combined data for each test.The critical value was computed to be $F_{crit}(2,74) = 1.64$ at a level of significance of $a = 0.2$ while the F-scores are shown in Table 2.

Table 2: F-scores for all six tests.

Piazza Garibalti - Piazza Vettovaglie	0.45
Piazza Garibalti - Piazza Gambacorti	0.05
Piazza Vettovaglie - Piazza Gambacorti	0.09
Largo Ciro Menoti – Piazza Garibalti	1.84
Largo Ciro Menoti – Piazza Vettovaglie	3.90
Largo Ciro Menoti – Piazza Gambacorti	1.80

The results suggest that the drift in Largo Ciro Menoti 's data was identified even between sensors placed 450 meters apart in the heart of the city. In addition, there was no rejection of the null hypothesis for the rest of the nodes regardless of the distance between them. This is attributed to the fact that the technique relies on the examination of the relation between the individual and combined regressions' SSR rather than point to point correlations between signals. As a result the method does not require high network density to ensure robust operation and produce accurate outcomes.

Case 2: Dissimilar variance between measurements

Comparisons between signals exhibiting similar variance like the ones demonstrated so far seem to provide accurate estimates on the existence, or non-existence, of stability issues in a monitoring network. What happens though when one signal exhibits much greater variance than the others? To answer this question we examined measurements coming from four locations presenting varying noise level patterns. In Figure 8 the increased variance in the measurements collected at Via Pietro Gori compared to the variance at all three other measurement positions is obvious.

The test was run from 22 January to 10 November 2014. In order to avoid interference of outliers with the outputs of the test extreme values were eliminated by setting a lower threshold at 42 dB and an upper limit as a function of the median of each dataset. The critical value was computed to be $F_{crit}(2,273)$ = 1.61 at a level of significance $a = 0.2$ and the results are summarized in Table 3.

Table 3: F-scores for the comparison between three public squares and one main street in the center of Pisa. The test was run from 22 January until 10 November 2014.

Piazza Garibalti – Piazza Omobono	2.63
Via Pietro Gori – Piazza Omobono	0.44
Via Pietro Gori – Piazza Garibalti	0.98
Piazza Vettovaglie - Piazza Omobono	1.91
Piazza Vettovaglie - Piazza Garibalti	0.03
Piazza Vettovaglie - Via Pietro Gori	0.75

According to the obtained F-ratios the sensors at Piazza Garibalti and Piazza Vettovaglie succeeded in identifying the drift. Nevertheless the second row of Table 3 suggests that the comparison between Via Pietro Gori and Piazza Omobono did not give any statistically significant indication of a fault in one of the sensors. This is due to increased variance in Via Pietro Gori's data caused by variability in human activity between different days of the week. Figure 8, shows that noise levels decrease significantly every Saturday and Sunday. This generates high sum of squared residuals when linear regression is applied on Via Pietro Gori's observations which do not increase significantly by the soft slope in Piazza Omobono's measurements when the two data sets are combined. The 'masking' is presented in Figure 9.

A work around would be the exclusion of all weekends from the data set. Doing so would certainly decrease the variance and could possibly make the data more suitable for long period comparisons. However there are other features that cannot be wiped out as easily and would impose errors in short term comparisons. These are seasonal characteristics, like the summer level dip discussed before, which is much more pronounced in noisier environments. However such features tend to exhibit similar behaviours at similar positions. Thus it is suggested that measurement locations are classified into soundscape categories and comparisons are made only between similar (but not necessarily near-by) locations. Over the last decade extensive research has taken place on soundscape classification [26] [27], [28], [29], [30] while various Sound Stabilization Time based methods can assist in this task as presented in [31], [32], [33].

Case 3: Abrupt changes in noise levels

Until now the detection of drifts within periods of one month or greater have been demonstrated. The principle of the proposed method however makes it suitable for detection of more severe anomalies which lead to abrupt changes in the measured noise levels. Such an example is presented in Figure 10.1 where a broken cable caused two weeks of faulty measurements for the Piazza Omobono station.
For this test an interval of thirteen days, from 31 July to 12 August 2014, was selected including only one day of faulty data. However missing measurements from Largo Ciro Menotti resulted in linear regressions applied on only six observations when this node was involved in the tests. The critical value for tests against Ciro Menoti was $F_{crit}(2,8) = 1.89$ at $a = 0.2$ level of significance while the rejection threshold was $F_{crit}(2,22) = 1.72$ for the remaining comparison at the same level of significance. In both comparisons the technique managed to detect the error; the results are shown in Table 4.

Table 4: Results of tests checking for abrupt faults in data collected between 31 July and 12 August 2014.

Ciro Menotti - Piazza Garibalti	0.12
Ciro Menotti – Piazza Omobono	1.92
Piazza Garibalti – Piazza Omobono	1.84

As the number of samples included in the test decreases, so do the degrees of freedom for the F test. This results in higher critical values and hence more lenient thresholds, assuming that the selected level of significance remains constant. However, the fewer the included samples are, the more susceptible the line fitting will be to extreme values; which makes it possible for the method to identify abrupt changes in the data when short time intervals are selected. Of course the sensitivity can be adjusted accordingly, by altering the level of significance, so that the test is tuned to identify such faults while effects caused by normal day to day variance in the measurements are minimized.

Case 4: Application of the Test on Data Coming from the Same Node

As discussed if no data scaling is applied the original form of the test can identify offset differences between data coming from one monitoring station but different time periods assuming that the soundscape is not expected to present significant level variations between them (e.g. same months but different years). An example of this application is demonstrated in Figure 11.

For a combined sample size of 45 the critical value was computed to be equal to $F_{crit}(2,41) = 1.67$ while the F score was 3.28 identifying correctly the approximately 2 dB drift which occurred during one year.

It must be noted though, that for this form of the test periods when seasonal effects cause very pronounced slopes in the time series, such as level dips during the summer months, should be excluded as they might result in wrong conclusions.

Case 5: A Useful Variation

Extracting the median night time SPL is a good way of reducing variance and making the data more suitable for application of the described method in order to test for calibration drifts or failures. Nonetheless, a variation of the test could aid in the exposure of other kinds of errors like noise floor level changes. This could be achieved by application of the proposed technique on L_{95} levels.

For the biggest part of the deployment L_{50} readings from Via Pietro Gori do not seem to present any particular issues apart from the first few weeks and maybe some isolated cases as seen in Figure 8. Figure 12, however, shows that the night time L_{95} levels demonstrate a major noise floor 'jump' from 35 dB(A) to about 39 dB(A). Examination of the measurement node after the deployment period was over showed that this noise floor increase occurred due to grounding issues of the unit. Application of the single node version of the test between periods 29 December to 2 February 2014 and 1 June to 30 June 2014 succeeded to identify this error achieving an F score of 4.54 when the critical value was $F_{crit}(2,60) = 1.65$ at $a = 0.2$. It is worth noting that in this case the selection of the time intervals for the comparison was not limited by any seasonal restraint because the noise floor levels are minimally affected by environmental characteristics.

Conclusions

Application of the described implementation of the Chow Test on noise levels measured at different locations can prove a very useful tool in the identification of stability issues in a noise monitoring network. It was demonstrated that high network density is not a prerequisite for this method to perform well as long as variance in the data is controlled. Routine measurements like night time median (L_{50}) levels can significantly reduce the data variance while it is believed that filtering and exclusion of weekend measurements could further aid towards this direction, if required. Obviously, the more monitoring stations exist the more accurate the stability assessment will be, however experimental results suggest that the proposed method is more susceptible to variance differences than low temporal correlations, hence the existence of great numbers of measurement units is not considered crucial.

Variations in the soundscapes' characteristics depending on population density and traffic can have an effect on the validity of the results especially when the existence of soft drifts is examined. For this reason classification of the different soundscapes and comparison between nodes found in similar noise environments instead of oversampling is proposed. In order to further improve accuracy the test can be tuned to the characteristics of each environment by adjusting the level of significance and choosing appropriate testing intervals. By doing so one could adjust the sensitivity and implement variations of the proposed method for the detection of other types of anomalies, such as abrupt failures, which happen in very short time intervals.

Apart from testing between nodes the discussed technique can be applied on signals coming from individual monitoring stations under the assumption that the noise environment locally does not change significantly with time. However, seasonal features can lead to wrong conclusions when only self-testing is used. Thus it is suggested to avoid the application of this form of the test during periods that exhibit seasonal peculiarities and when self-testing is used always to cross check the results with those obtained by the comparison between multiple nodes.

Finally, the method is suitable for application on other parameters, statistical- and otherwise. An example of the test run on L_{95} levels showed that the same technique can be used to expose noise floor shifts. In a similar way the proposed technique can be expanded and applied on the response of any group of linearly behaving elements to help detect faults which might not be easily identified otherwise. It has been demonstrated that night time L_{50} exhibits less variance than L_{eq} and provides more suitable time series for the application of the test in order to identify slow drifts from calibration or abrupt changes in the response of the system. However, statistical levels only present a portion of the true behaviour of the sound environment. Examining only the median might lead to incapability to identify errors related to the upper or lower end of the level range at a given location. Hence it is recommended that other parameters such as L_{eq} or L_{10} and L_{90} are also examined using suitable methods and the results of all tests are combined in order to cover all possible failure scenarios.

Acknowledgements

This project has been funded by National Measurement Office of the United Kingdom Department for Business, Innovation and Skills. The data used to demonstrate the technique were made available by Dustlab and were collected as a part of 'SENSEable Pisa' project.

References

[1] G. Andriaens, "Smart Cities, Smart Europe: Putting Our Energy into Inovation and Sustainability," Brussels, 2014.

[2] M. Becker, S. Caminit, D. Fiorella, L. Francis, P. Gravino, M.M. Haklay, A. Hotho, V. Loreto, J. Mueller, F. Ricchiuti, V. D. P.Servedio, A. Sîrbu, F. Tria, "Awareness and Learning in Participatory Noise Sensing," *PLOS One,* vol. 8, no. 12, 2013.

[3] C.A. Kardous, P.B. Shaw., "Evaluation of smartphone sound measurement applications.," *L.Acoust.Soc.Am.,* p. 135(4), April,2014.

[4] N. Maisonneuve, N. Matthias , "Participatory noise pollution monitoring using mobile phones.," *Inf. Polity 15,* pp. 51-71, 2010.

[5] Timothy Van Renterghem, Pieter Thomas, Frederico Dominguez, Samuel Dauwe, Abdellah Touhafi, "On the ability of consumer electronics microphones for environmental noise," *J. Environ. Monit,* p. 544, 2011.

[6] R. Barham, et.al., "Development and performance of a multi-point distributed environmental noise measurement system using MEMS microphones.," in *Euronoise 2009*, Edinbrough, 2009.

[7] G. Memoli, D.Dawson, R.Barham et al., "Towards the acoustical characterization of an Intensive Care Unit.," *Applied Acoustics 79,* pp. 124-130, 2014.

[8] Richard Barham, M. Chan, M. Cand, "Practical experience in noise mapping with MEMS microphone based distributed noise measurement system.," in *InterNoise 2010*, Lisbon, 2010.

[9] R. Jackett, R. Barham, "Development of the Minim Outdoor Noise Monitoring System, Report AC9,," NPL, London, 2014.

[10] R. Barham, "New Paradigm for Environmental Noise Measurement," Teddington, 2014.

[11] V. Chandola, A. Banerjee, V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.,41,* pp. 15:1-15:58, 2009.

[12] Y. Zhang, N. Merantia, P. Havinga, "Outlier detection techniques for wireless sensor networks: A survey.," *Communications Surveys and tutorials, IEEE,12,* pp. (2):159-170, 2010.

[13] S. Dauwe, D. Oldoni,B. De Baets,T. Van Renterghem, D. Botteldooren, B Dhoedt, "Multi-Criteria Anomaly Detection in Urban Noise Sensor Networks," *Environmental Science : Processes & Impacts,* vol. 16, no. 10, pp. 2249-2258, 2014.

[14] V. Bychkovsky, S. Megerian, D. Erstin, M. Potkonjak, "A collaborative approach to in-place sensor calibration.," in *In Proceedings of the Second International Workshop on Information Processing in Sensor Networks*, 2003.

[15] L. Balzano, R. Nowak, "Blind Calibration of Networks of Sensors: Theory and Algorithms," in *Networked Sensing Information and Control*, Springer, 2007, pp. 9-37.

[16] R. Tan, G. Xing, Z.i Yuan, X. Liu, J. Yao, "System-level calibration for data fusion in wireless sensor networks," *ACM Transactions on Sensor Networks vol9,* 2013.

[17] G. C.Chow, "Tests of equality between sets of coefficients in two linear regression.," *Econometrica,* vol. 3, no. 26, pp. 591-605, 1960.

[18] F.M. Fisher, “Tests of equality between two sets of coefficients in two linear regression: An expository note,” *Econometrica,* vol. 38, no. 2, pp. 361-366, 1970.

[19] T.D. Cook, D.T. Campbell, Quasi-experimentation: Design & analysis issues for field settings., Chicago: Rand McNally, 1979.

[20] W.P. Shadish, T. D. Cook, D. T. Cook, Experimental and quasi-experimental designs for generalized causal inference., New York: Houghton Mifflin Company, 2001.

[21] I. H. Zuckerman, E. Lee, A.K. Wutoh, Z.Xue, B. Stuart, “Application of regression-discontinuity analysis in pharmaceutical health services research,” *Health Service Research,* vol. 41, no. 2, pp. 550-563, 2006.

[22] J.P. Braden, T.J. Bryant, “Regression discontinuity designs: Applications for school psychology,” *School Psychology Review,* vol. 19, no. 2, pp. 232-239, 1990.

[23] H.B. Lee, “Using the Chow Test to Analyze Regression Discontinuities,” *Tutorials in Quantitative Methods for Psychology,* vol. 4, no. 2, pp. 46-50, 2008.

[24] Kostas Sotirakopoulos, Richard Barham, Ben Piper, Richard Jacket, “On The Use of Linear Regression for the Assessment of Stability in Noise Monitoring Networks: A Practical Example,” in *Proceedings of Euronoise 2015*, Maastricht, 2015.

[25] L.Nencini, P. De Rosa, E. Ascari, B. Vinci, N. Alexeeva, “SENSEable Pisa: A wireless sensor network for real-time noise mapping,” in *Proceedings of Euronoise 2012*, Prague, 2012.

[26] Chu, S. Narayanan, S. Kuo, C.-C.J., “Environmental Sound Recognition With Time–Frequency Audio Features,” *AUDIO, SPEECH, AND LANGUAGE PROCESSING,* vol. 17, no. 6, pp. 1142-1158, 2009.

[27] Jin Yong Jeon, Joo Young Hong, “Classification of urban park soundscapes through perceptions of the acoustical environments,” *Landscape and Urban Planning ,* vol. 141, pp. 100-111, 2015.

[28] Ling MA, B.Milner, D. Smith, “Unsupervised Feature Learning For Uran Sound Classification,” *ACM Transactions on Speech and Language Processing,* vol. 3, no. 2, pp. 1-22, 2006.

[29] Barbara Lebiedowska, “Acoustic background and transport noise in urbanised areas: A note on the relative classification of the city soundscape,” *Transportation Research Part D: Transport and Environment,* vol. 10, no. 4, pp. 341-345, 2005.

[30] Manon Raimbault, Danièle Dubois, “Urban soundscapes: Experiences and knowledge,” *Cities,* vol. 22, no. 5, pp. 339-350, 2005.

[31] Torija, A. J., Ruiz, D. P., & Ramos-Ridao, A., “Estimation procedure of the descriptor LAeq, T from the stabilization time of the sound pressure level value,” *Noise and Vibration Worldwide,* vol. 43, no. 1, p. 11–18, 2012.

[32] Prieto, C. Barrigon, JM, “Stabilisation patterns of hourly urban sound levels,” *Environmental Monitoring and Assessment,* vol. 187, no. 1, 2015.

[33] Brocolini, L., Lavandier, C., Quoy, M., & Ribeiro, C., “Measurements of acoustic environments for urban soundscapes: choice of homogeneous periods, optimization of durations, and selection of indicators,” *Journal of the Acoustical Society of America,* vol. 134, no. 1, p. 813–821, 2013.

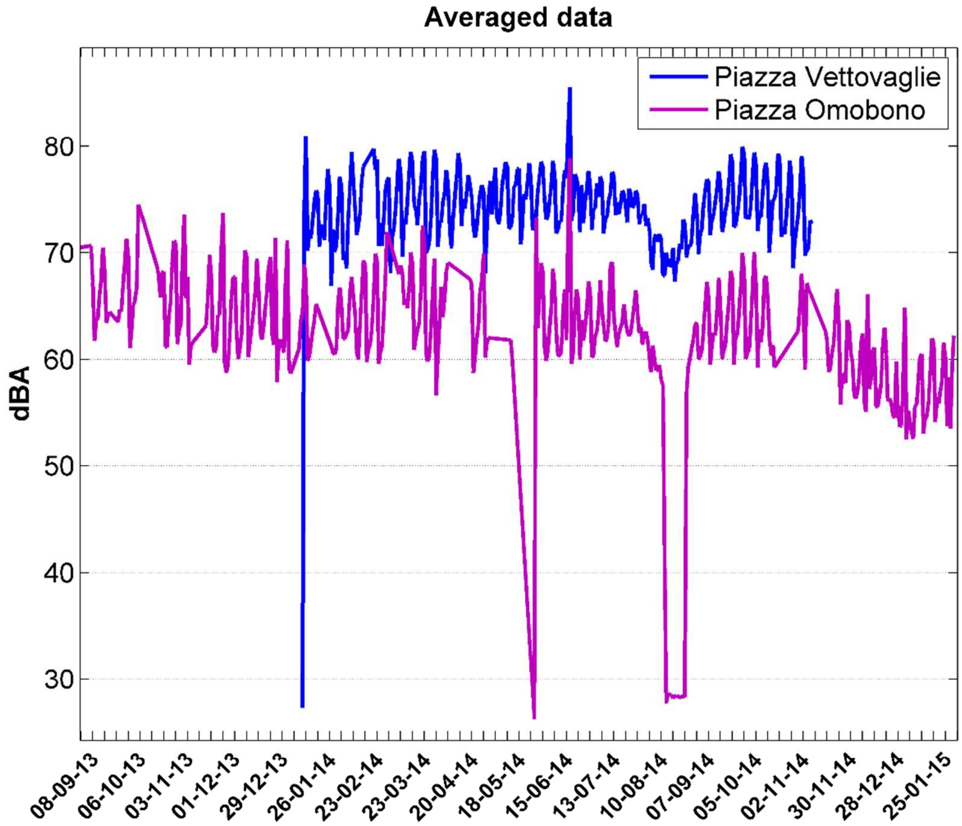


Figure 1: Daily averaged broadband Leq levels measured at Piazza Vettovaglie and Piazza Omobono in Pisa.

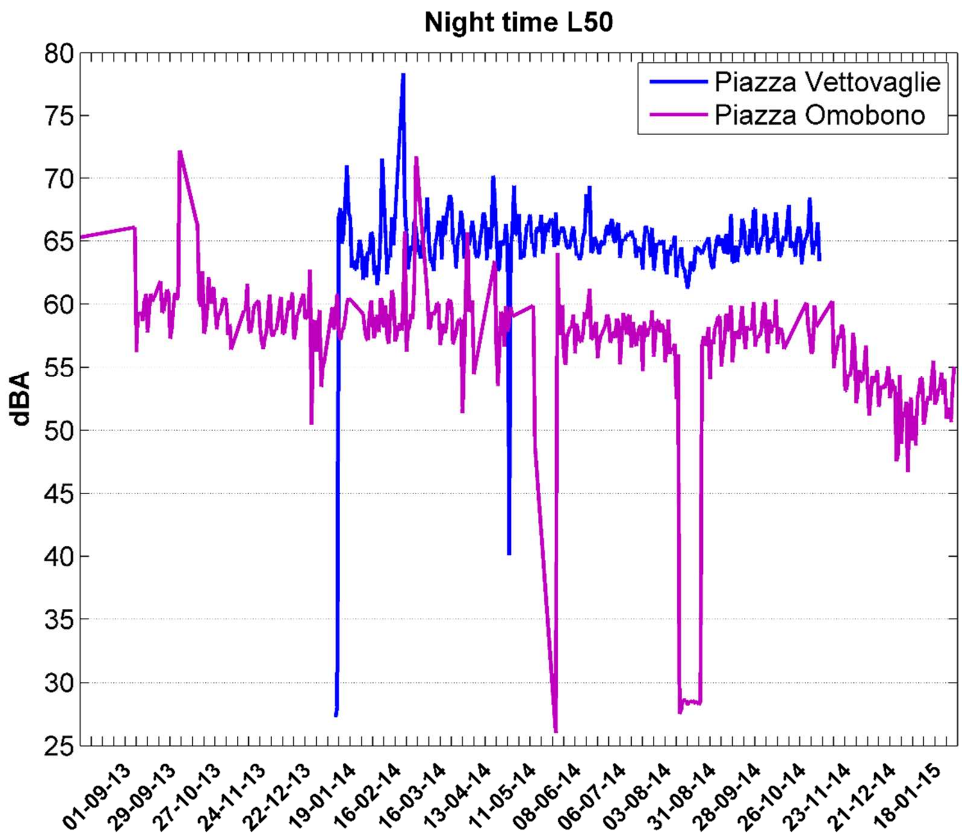


Figure 2: Night time L_{50} levels measured at Piazza Vettovaglie and Piazza Omobono in Pisa.

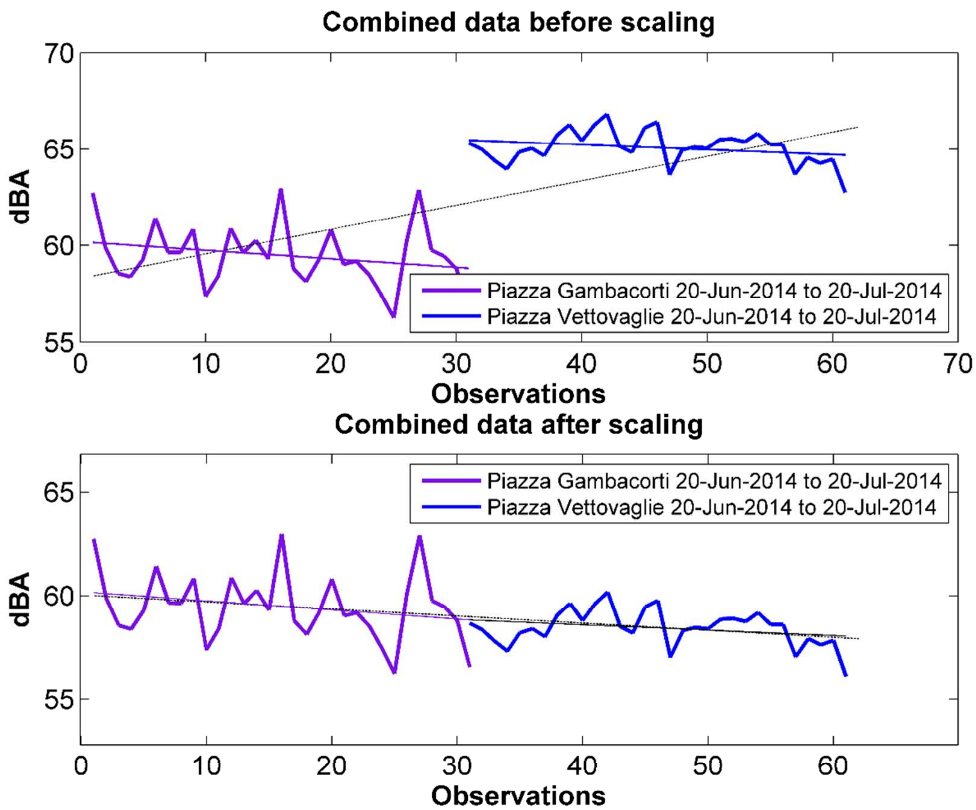


Figure 3.1: Combined L_{50} data and their regression lines before scaling.
Figure 3.2: Combined L_{50} data and their individual regression lines after scaling is applied.

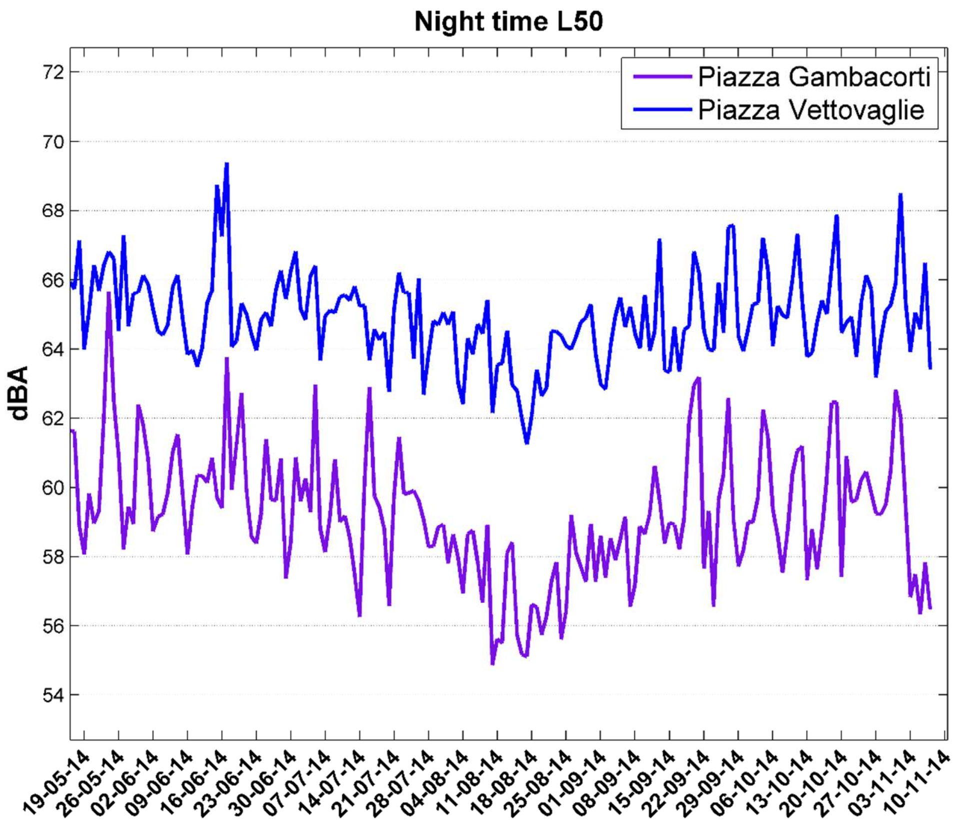


Figure 4: Night time L_{50} levels for the whole duration of the deployment.

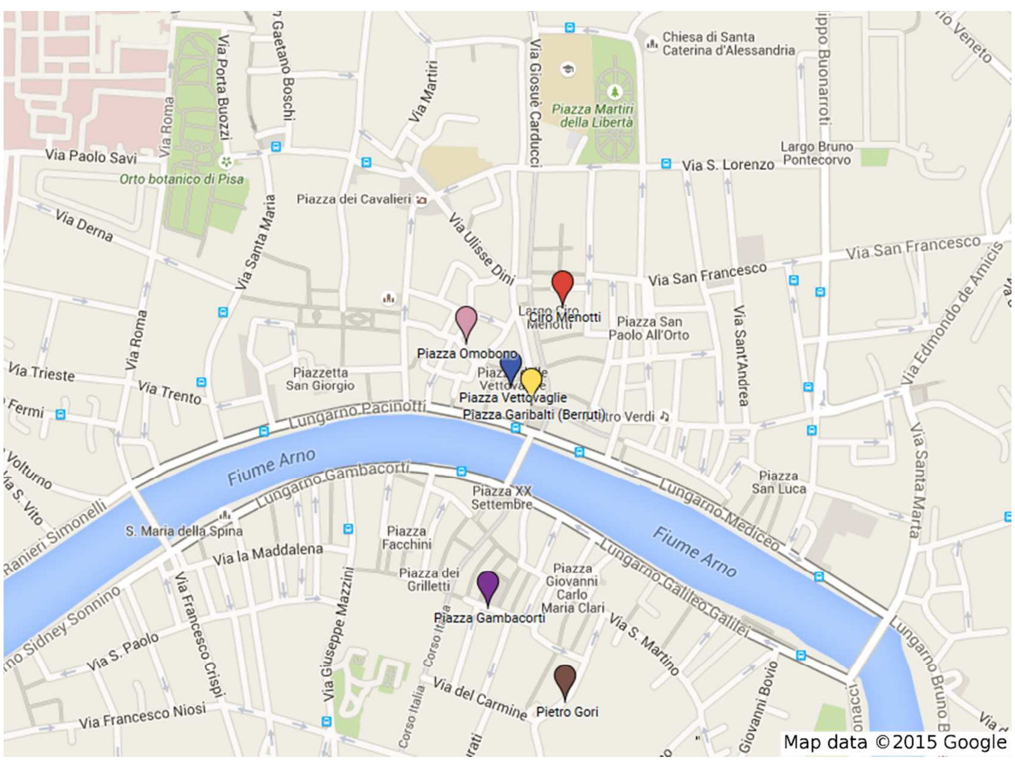


Figure 5: Map showing the location of the measurement points in the centre of Pisa.

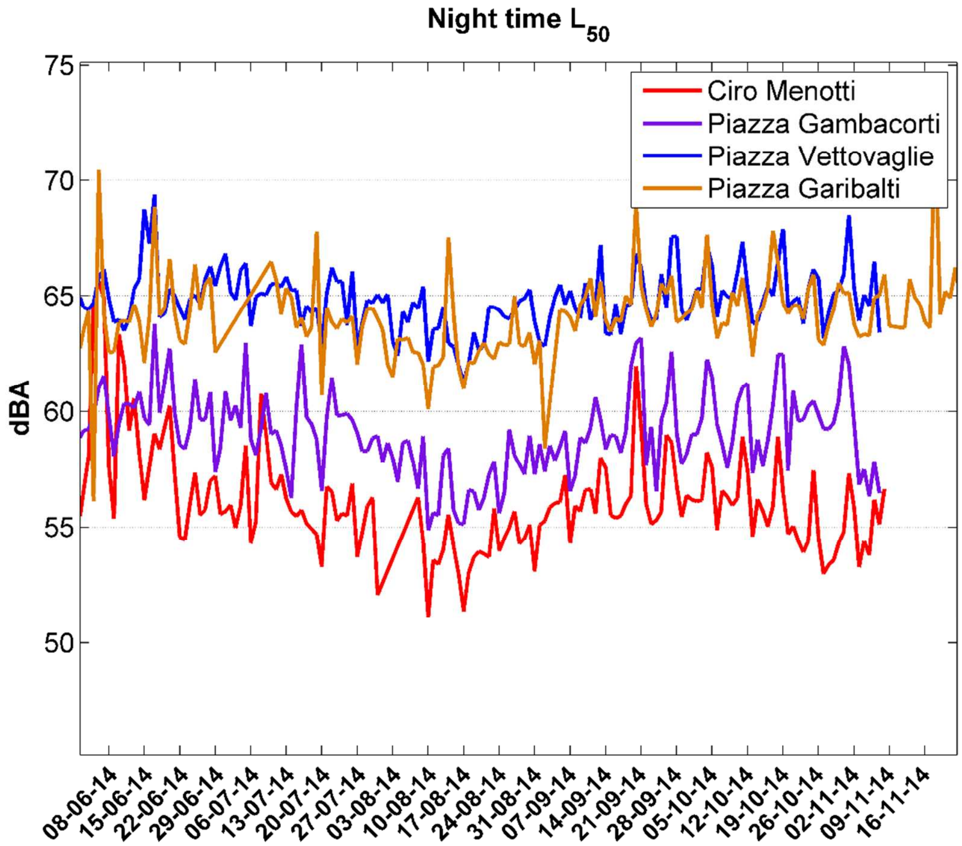


Figure 6: Night time L_{50} measurements at four different locations in Pisa.

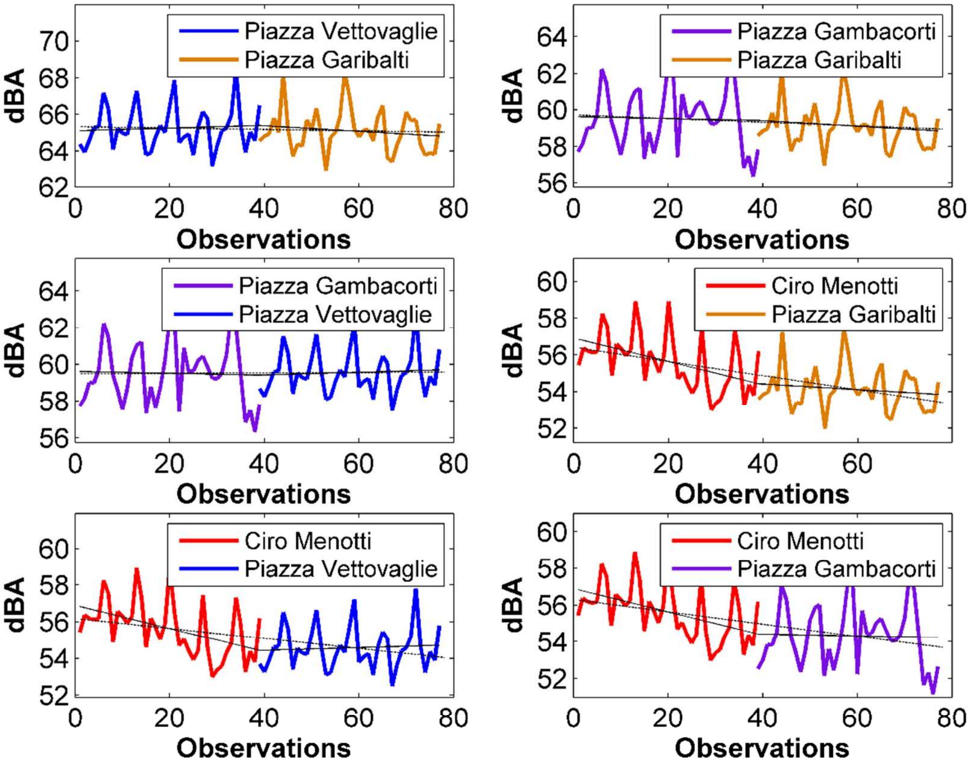


Figure 7: Comparison of fitted lines between individual and combined data sets.

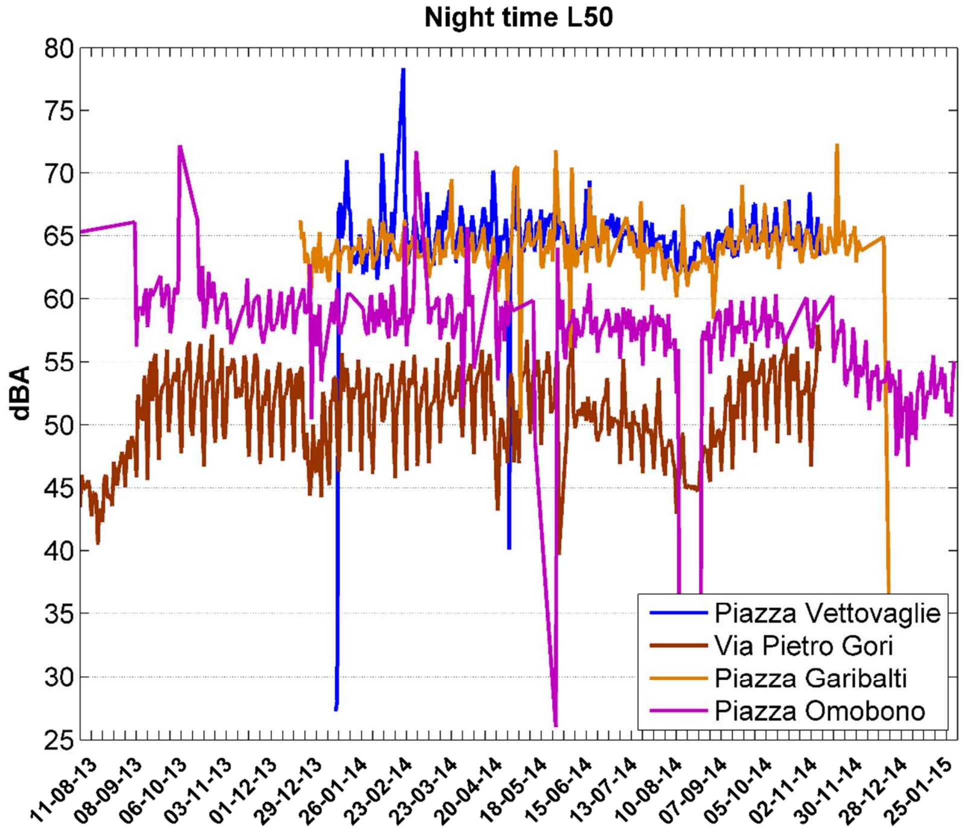


Figure 8: Time series of night time L_{50} measurements at four different locations in Pisa.

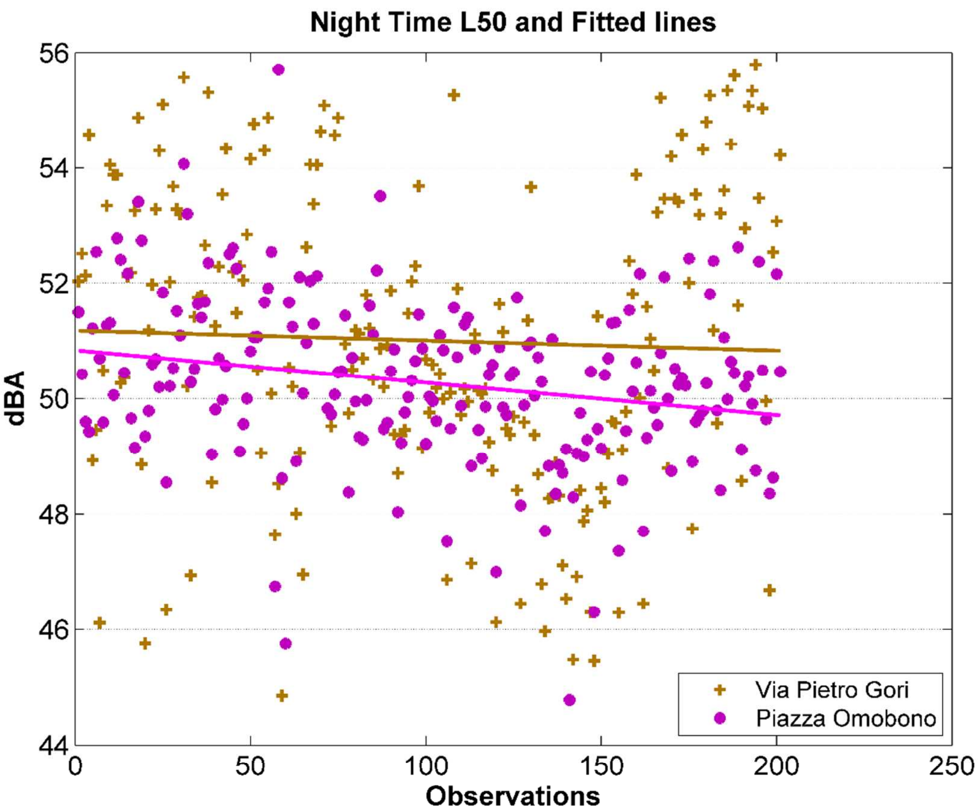


Figure 9: Night time L_{50} levels measured at Via Pietro Gori and Piazza Omobono between 22 January and 10 November 2014 and their fitted lines. The much greater variance of Via Pietro Gori's measurements masks the deviation from normal operation at Piazza Omobon

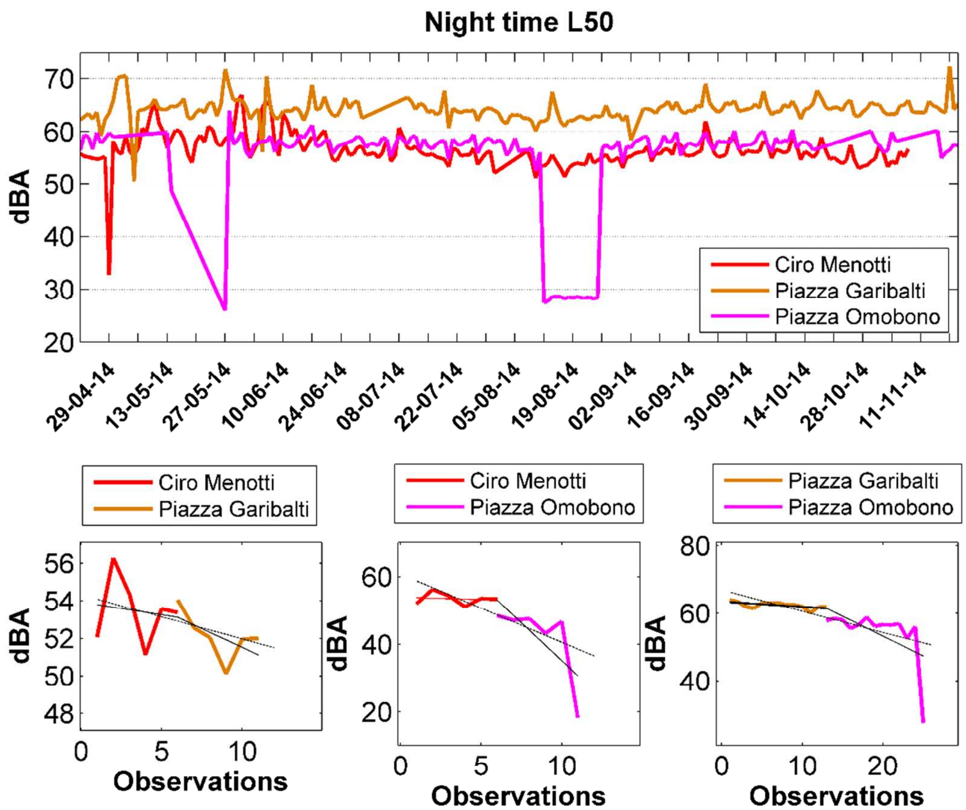


Figure 10.1: Night time L_{50} levels measured at Ciro Menoti, Piazza Garibalti and Piazza Omobono

Figure 10.2-4: Comparison between individual and combined linear regressions on data collected from 31 July to 12 August 2014

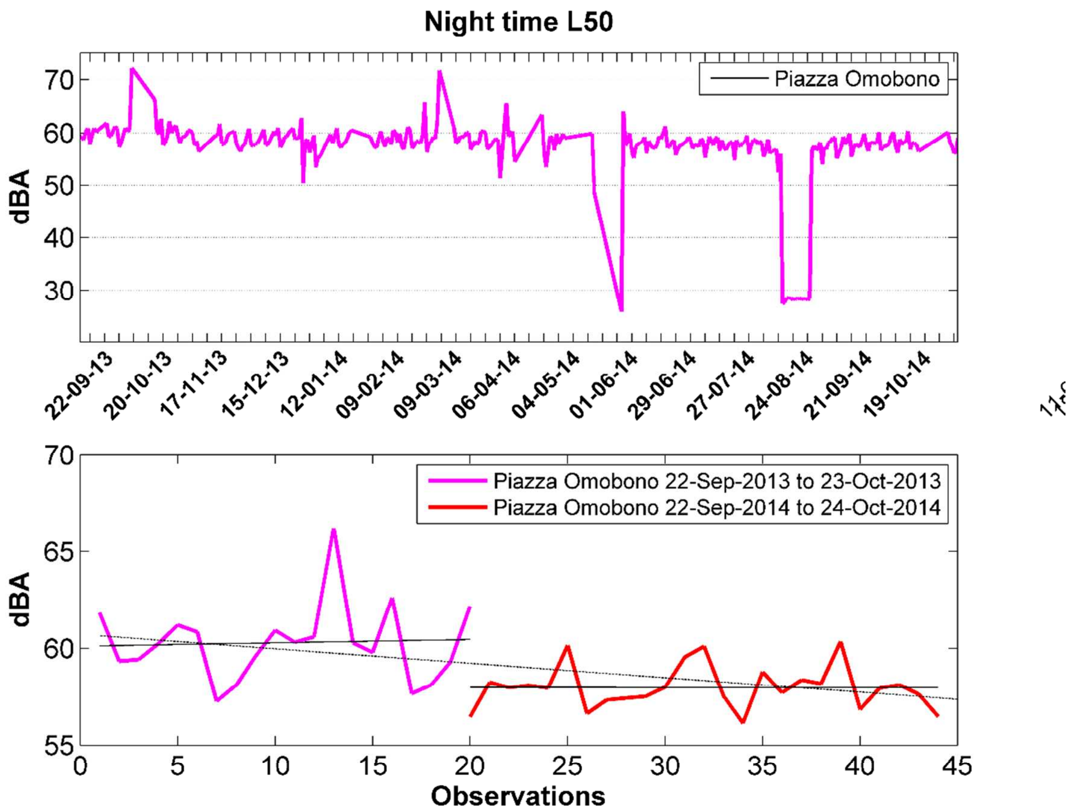


Figure 11.1: Night time L_{50} measured at Piazza Omobono.
Figure 11.2: Comparison of the individual and combined regressions applied on the time intervals 22 September to 23 October 2013 and 22 September to 24 October 2014.

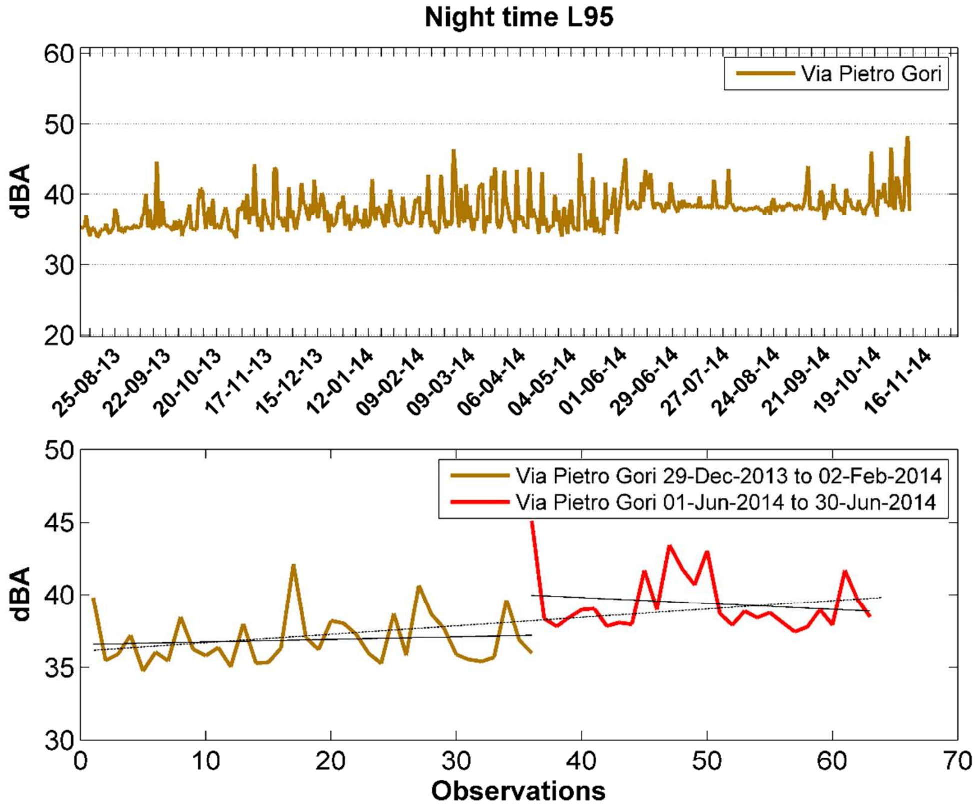


Figure 12.1: Night time L_{95} measurement at Via Pietro Gori.
Figure 12.2: Comparison between individual and combined data regression lines for the time periods 29 December 2013 to 2 February 2014 and 1 June to 30 June 2014.